MINISTRY OF FISHERIES
Te Tautiaki i nga tini a Tangaroa

ISSN 1175-1584

Catch-per-unit-effort (CPUE) analysis of hake (*Merluccius australis*) for Sub-Antarctic and Chatham Rise fisheries from 1989–90 to 2000–01

N. L. Phillips
M. E. Livingston

New Zealand Fisheries Assessment Report 2004/19
April 2004

Catch-per-unit-effort (CPUE) analysis of hake (*Merluccius australis*) for Sub-Antarctic and Chatham Rise fisheries from 1989–90 to 2000–01

N. L. Phillips
M. E. Livingston

NIWA
Private Bag 14901
Wellington

New Zealand Fisheries Assessment Report 2004/19
April 2004

Published by Ministry of Fisheries
Wellington
2004

ISSN 1175-1584

Citation:
Phillips, N.L.; Livingston, M.E. (2004).
Catch-per-unit-effort (CPUE) analysis of hake (*Merluccius australis*) for Sub-Antarctic and Chatham Rise fisheries from 1989–90 to 2000–01.
New Zealand Fisheries Assessment Report 2004/19. 39 p.

This series continues the informal
New Zealand Fisheries Assessment Research Document series
which ceased at the end of 1999.

EXECUTIVE SUMMARY

Phillips, N.L.; Livingston, M.E. (2004). Catch-per-unit-effort (CPUE) analysis of hake (*Merluccius australis*) for Sub-Antarctic and Chatham Rise fisheries from 1989–90 to 2000–01.

New Zealand Fisheries Assessment Report 2004/19. 39 p.

This report provides estimates of standardised catch-per-unit-effort (CPUE) indices for the Chatham Rise and Sub-Antarctic hake fisheries from 1989–90 to 2000–01.

Commercial catch and effort data from Trawl Catch Effort Processing Returns (TCEPR forms) from 1989–90 to 2000–01 were extracted from the Ministry of Fisheries catch and effort database. The data were extensively groomed to correct errors and to remove tows that were believed to be misreported.

Standardised CPUE indices were calculated using lognormal generalised linear models as used in previous years and the application of mixed-effect models was evaluated. CPUE indices using the two methods for a range of data subsets were calculated for several areas (Chatham Rise, Statistical Area 404, Sub-Antarctic) using core vessels (i.e., vessels that caught at least 80% of the hake catch and have operated in the fishery for three or more consecutive years).

CPUE indices derived from the Chatham Rise fishery fluctuated from 1989–90 to 1995–96 with no trend, declined from 1995–96 to 1998–99, and increased in 1999–2000 and 2000–01. The indices derived from Statistical Area 404 showed a declining trend, apart from 2000–01 when there was a significant increase. CPUE indices derived from the Sub-Antarctic fishery showed a declining trend, apart from the 1995–96 fishing year when there was an increase.

The CPUE indices derived from the Chatham Rise fishery as a whole did not correlate well with trawl survey abundance indices, but CPUE indices from the Statistical Area 404 fishery within the Chatham Rise showed a reasonable correlation. It is, therefore, unclear how well CPUE reflects hake abundance in the Chatham Rise. The Sub-Antarctic CPUE indices showed declining trends for each vessel subset, and correlated reasonably well with trawl survey abundance indices, suggesting that CPUE may reflect abundance in this area.

The application of mixed-effect models in CPUE analysis showed some improvements over the generalised linear models. The model selected fewer categorical variables, and the CPUE indices were less prone to bias when using unbalanced data sets. The trends were similar to the indices derived from more conventional methods. Further research on variable selection and the inclusion of tows with zero catches may improve the model selection.

1. INTRODUCTION

Hake are widely distributed throughout 250–800 m depths south of latitude 40° S in the New Zealand Exclusive Economic Zone (EEZ) (Colman 1995). The fishery is managed as three main areas, or Quota Management Areas (QMA); the Challenger Plateau and west coast of the South Island (HAK 7), the eastern Chatham Rise (HAK 4), and the remainder of the EEZ comprising the Auckland, Central, Southeast (Coast), and Sub-Antarctic (HAK 1) (Figure 1). An administrative stock has been established for the Kermadec area (HAK 10) but no catch of hake has been recorded from that area.

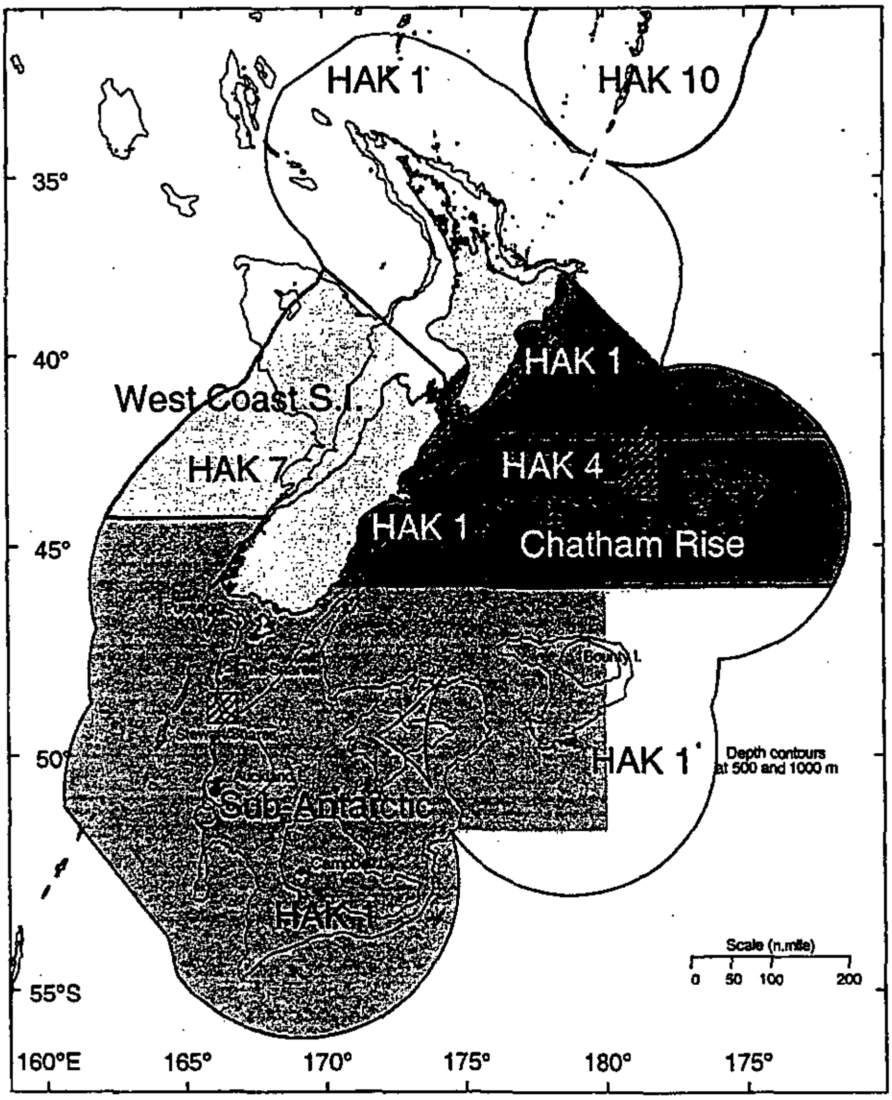

Figure 1: Quota Management Areas (QMA), names of locations used in this report, and proposed fisheries (shaded areas) for hake in the New Zealand EEZ. Statistical Area 404 spawning area is shown as the hashed region in HAK 4, and the Stewart/Snares spawning area is shown as the hashed region in the Sub-Antarctic.

The largest fishery is HAK 7, where landings in recent years have been about 7000 t, which is higher than the TACC (Annala et al. 2003). In HAK 1, most of the catch is taken from the Sub-Antarctic, and landings have been just under 4 000 t in recent years. Landings from HAK 4 have declined from about 3500 t (1995–96) to less than 1500 t (Annala et al. 2003).

The current stock hypothesis suggests that there are three separate hake stocks (or fisheries) (Colman 1988), namely the west coast South Island stock (HAK 7), the Sub-Antarctic stock (the area of HAK 1 that encompasses the Sub-Antarctic), and the Chatham Rise stock (HAK 4 and the area of HAK 1 on the western Chatham Rise and the east coast North Island) (Figure 1).

In this study, data from the Chatham Rise and Sub-Antarctic fisheries were analysed. In these fisheries, hake are caught mainly as bycatch by trawlers targeting hoki (80% of tows in Sub-Antarctic, 78% on the Chatham Rise), but some targeting occurs, particularly in Statistical Area 404 on the Chatham Rise, and in the channel that lies between the Stewart and Snares Shelf and Auckland Shelf in the Sub-Antarctic (Stewart/Snares box) (Figure 1). About half (51%) of the hake catch from the Chatham Rise is targeted in Statistical Area 404 during spawning between September and January (Phillips 2003). The remainder of the hake catch occurs while targeting hoki (35%) and other species throughout the year. Forty percent of the hake catch from the Sub-Antarctic is taken as target species in the Stewart/Snares box spawning area from September to January (Phillips unpublished results).

Commercial catch and effort data have previously been analysed to produce CPUE indices for HAK 7 (Colman et al. 1991, Vignaux 1992), for HAK 1, 4, and 7 (Kendrick 1998), and for HAK 4 (Dunn et al. 2000b) using the methodology of Gavaris (1980) and Vignaux (1994). An analysis was completed in 2001, but was not published due to a problem with misreporting of hake catches taken from HAK 7 that were declared as having been taken from Chatham Rise. The misreported tows have been identified and removed in the present study. For details of the methods for identifying the misreported tows, see Dunn (2003).

This study differs from previous studies by obtaining abundance indices using mixed-effect models as well as the commonly used generalised linear models. A brief description of the use of mixed-effect models is presented in Section 2.2; but more details were given by Pinheiro & Bates (2000).

The advantage of using a mixed-effect model is that the number of categorical variables can be reduced by allowing them to be random variables rather than fixed. This greatly improves the statistical properties of the model; for example, the method prevents model overfitting (Richardson & Welsh 1995, Barrowman 2000).

This report fulfils Objective 1 (To update the analysis of the commercial catch and effort data for HAK 1 and 4 with the inclusion of data up to the end of the 2000/2001 fishing year) of Ministry of Fisheries project HAK 2001/01.

2. METHODS

2.1 Data selection

The data comprise commercial catch and effort data from all vessels catching and/or targeting hake on the Chatham Rise and Sub-Antarctic fisheries from 1989–90 to 2000–01. All misreported tows and suspected tows, as identified by Dunn (2003), were removed from the data set

The TCEPR data extracted from both fisheries included information on where vessels fished, the top five species caught, how much hake was caught, and vessel characteristics. Table 1 lists the variables extracted and the derived variables used in the CPUE analysis.

2.1.1 Variables

Fishing year was a 12-level categorical variable covering October–September each year. The 1992 fishing year therefore refers to 1991–92 and so on for subsequent fishing years.

The variable *fishing method* was a 3-level categorical variable consisting of bottom trawl, midwater trawl, and a new method (definition midwater/bottom) where fishers use a midwater trawl, but trawl on the seafloor with a reduced headline height.

Tow distance was deemed the best measure of effort for this study because it incorporated temporal changes in *tow duration* and *vessel speed. Tow distance* was calculated as a product of *tow duration* and *vessel speed.*

Subareas for each stock were based on subjective assessment of the geographical distribution of fishing effort, the bathymetry of the area, and area trends in catch rate. In determining the number and extent of the *Subareas* within each stock, it was necessary to balance the spatial scale of the *Subareas* to encompass any significant areal differences in catch rate, while ensuring each *subarea* included a sufficient quantity of tow records. Tows were allocated to *Subareas* based on the start position of the tow. Definitions of the *Subareas* are given in Appendix A, Figure A1 for the Chatham Rise and Figure A2 for the Sub-Antarctic.

Table 1: Description of variables used in the CPUE analysis. (*, derived variable.)

Variable	Type	Description
Year	Categorical	Fishing year (October-September the following year)
Vessel	Categorical	Encrypted vessel identification number
Start date	Continuous	Date at the start of the tow
Start longitude	Continuous	Longitude at the start of the tow
Start latitude	Continuous	Latitude at the start of the tow
Fishing method	Categorical	Method of trawl
Wingspread	Continuous	Wingspread (m) of the net at the start of the tow
Headline height	Continuous	Headline height (m) of the net at the start of the tow
Bottom depth	Continuous	Bottom depth (m) at the start of the tow
Groundrope depth	Continuous	Ground rope depth (m) at the start of the tow
Target species	Categorical	Species of fish targeted
Total catch	Continuous	Estimated catch (t) of target and bycatch species for the tow
Hake catch	Continuous	Estimated catch of hake (t)
Vessel speed	Continuous	Speed of vessel (kn) during tow
*Net off bottom**	Continuous	Height of the net from the bottom (m)
*Tow duration**	Continuous	Duration of the tow (hrs)
*Tow distance**	Continuous	The distance of the tow (n. miles)
*Fishing day**	Continuous	Number of days since the start of the fishing year
*Month**	Continuous	Month of the year
*SOI**	Continuous	Mean Southern Oscillation Index for the month of tow
*Vessel experience**	Continuous	Number of years the vessel has been operating in the fishery
*Subarea**	Categorical	Subarea within fishstock
*CPUE**	Continuous	Catch-per-unit-effort (catch per tow)

2.2 Models used

Estimates of relative year effects were obtained from the more conventional method of generalised linear models (GLM) with fixed variables only, and a mixed-effect model, in which data were modelled using both fixed and random variables.

2.2.1 Linear models

A typical linear (fixed) model used in previous CPUE analyses of hake is normally written as

$$Y_{ij} = \beta i + \varepsilon_{ij}$$

where

Y_{ij} is the value of the response variable of *j*th individual in the ith group
βi is the mean of the *i*th group
ε_{ij} is a random error term $\sim N(0,\sigma^2)$

Even though the fixed-effects model accounts for the group effects, it does not provide a useful representation of the data. The problem is that it models only the specific sample of effect of each vessel group used, but the main interest is in the population from which the sample was drawn. The fixed model does not provide an estimate of the between group variability, and the number of parameters in the model increases linearly with the number of groups.

2.2.2 Mixed-effect models

Mixed-effect models incorporate both fixed and random effects. Fixed effects can be considered as the effects attributable to a finite set of levels of factors that occur in the data. For example, the year effects are fixed, because there is only a finite number of fishing years (i.e., 1989–2001). Random effects are usually attributable to an infinite set of levels of a particular factor. For example, there are, in theory, an infinite number of vessels that could fish on a population of hake. If this was true, we could assume that the actual vessels used in the fishery are a random sample of all possible vessels. By treating a variable as random, it is possible (and desirable) to estimate the variance components of the random sample, which in this case would be vessel variance and error variance (Searle 1987).

The mixed-effect model is written as

$$Y_{ij} = \beta + b_i + \varepsilon_{ij}$$

$$\beta = \frac{\sum_{i=1}^{m} \beta_i}{m}$$

$$b_i = \beta_i - \beta$$

where

β is the mean response of the fixed effects (e.g., year)
b_i is a random variable (e.g., vessel) representing the deviation $\sim N(0,\sigma_b^2)$
m is the number of groups
ε_{ij} is a random error term $\sim N(0,\sigma^2)$

2.3 Standardised CPUE analysis

Fishing year was treated as a categorical value so that the regression coefficients of each year could be estimated independently. The relative year effects calculated from the regression coefficients represent the change in CPUE from year to year, all other variables having been taken into account. It therefore potentially represents an index of relative abundance.

The fixed variables for both the mixed-effect and lognormal GLM models were selected by a forward stepwise multiple regression-fitting algorithm (Chambers & Hastie 1991, Venables & Ripley 1994). The algorithm generates the final regression model iteratively by adding each term that results in the greatest reduction in residual deviance (greater than 1%) with each run. The stepwise algorithm also considers first order interaction terms. At each step, all first order interactions between variables selected up to that point are evaluated. As earlier, terms that resulted in a reduction of more than 1% in residual deviance were added to the model, or else excluded. As the primary interest in the model was an estimate of relative year effects, possible interactions with *year* were not evaluated.

To compare results of the mixed-effect models and lognormal models, the random effects in the mixed model were *method* nested within *vessel*.

Vessel effects were incorporated into the CPUE standardisation to allow for likely differences in fishing power between vessels. Vessels not involved in the fishery for consecutive years, or that have participated for only a couple of years, provide little information for the standardisations (Knuckey et al., CSIRO) and can result in model over-fitting (Francis 2001). CPUE analysis was investigated for "core" vessels that caught 80% of the hake catch and were involved in the fishery for at least 3 consecutive years. Figures B1–B4 (Appendix B) show the number of tows by vessel for each fishing year.

Catch-per-tow was chosen as the measure of CPUE because catch per nautical mile does not increase in a linear fashion (Figure 2). For example, there are high catch rates for short tows and low catch rates for long tows, so the length of the tow in time or distance is not a relevant measure of effort.

For this study, tows from four different areas were considered, as follows.

- Chatham Rise fishery (including Statistical Area 404):
  -Mixed-effect model of catch rates for core vessels.

- Chatham Rise fishery (excluding Statistical Area 404)
  -Mixed-effect model of catch rates for core vessels.
  -Lognormal model of catch rates for core vessels.

- Statistical Area 404 (Chatham Rise) fishery:
  -Mixed-effect model of catch rates for core vessels.
  - Lognormal model of catch rates for core vessels.

- Sub-Antarctic fishery:
  -Mixed-effect model of catch rates for core vessels.
  - Lognormal model of catch rates for core vessels.

- Stewart/Snares (Sub-Antarctic) fishery:
  -Mixed-effect model of catch rates for core vessels.

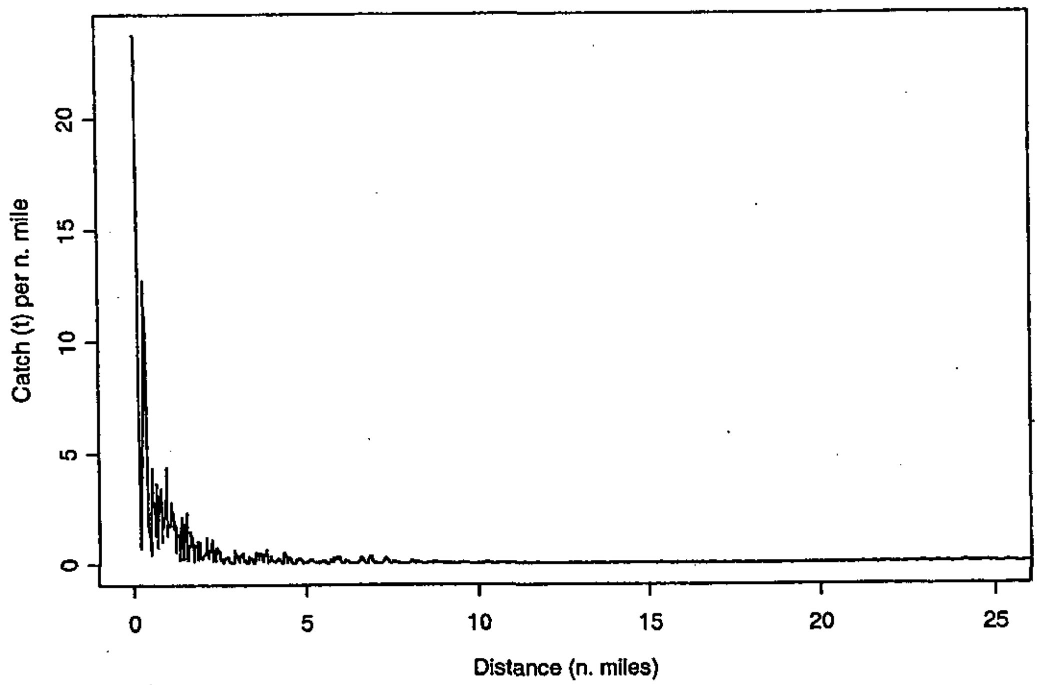

Figure 2: Relationship between hake catch rate (t/n. mile) and length of the tow. The data have been smoothed using a kernel smoother (Watson 1966).

3. RESULTS

3.1 Chatham Rise fishery (including Statistical Area 404)

The Chatham Rise fishery was fished by a total of 171 unique vessels catching 17 046 t of hake (Table 2); There were 37 957 records. Restricting the data to core vessels resulted in a 46% reduction in records to 20 595 records and a 22% reduction in hake catch to 13 373 t (28 unique vessels).

Table 2: Summary of the number of records, catch and number of unique vessels for the Chatham rise fishery by fishing year.

	All vessels			Core vessels		
Fishing year	No. records	Catch (t)	No. vessels	No. records	Catch (t)	No. vessels
1990	952	699	21	446	528	6
1991	2 265	702	35	1 567	458	10
1992	2 897	1 370	43	1 707	967	14
1993	2 792	1 167	47	1 666	803	17
1994	1 784	1 060	40	702	925	12
1995	2 863	1 708	52	1 709	1 286	15
1996	3 313	2 567	68	1 960	2 316	15
1997	4 476	1 879	71	2 455	1 529	16
1998	4 796	1 261	67	2 402	851	20
1999	4 361	1 646	69	2 224	1 338	19
2000	3 590	950	54	1 926	726	18
2001	3 868	2 037	58	1 831	1 645	18
Total	37 957	17 046	171	20 595	13 373	28

The variables selected by the stepwise regression are given in Table 3. Four variables were selected with a total r^2 (percentage of residual deviance explained) of 58%, with *subarea* explaining 51% of the residual deviance. The other variables selected (in order) were *target species*, *target species**subarea*, and *fishing year*.

Table 3: Variables in order of selection and the percentage of the deviance explained (r^2) for the fixed effects in the mixed-effect model for the Chatham Rise fishery (includes Statistical Area 404). (*, interaction terms).

Variable	r^2
Subarea	51.0
Target species	53.8
*Target species*Subarea*	55.8
Fishing year	57.6

The resulting year effects using the mixed-effect model are listed in Table 4 and plotted in Figure 3. The indices fluctuated from 1989–90 to 1995–96, decreased for the next 3 years, followed by an increase in 1999–2000, and remained at a similar level for the 2000–01 fishing year.

Table 4: Relative year effects and 95% confidence intervals (CI) by fishing year using the Mixed-effect model for the Chatham Rise fishery (includes Statistical Area 404).

Fishing year	Index	95% C.I.
1990	1.10	1.02–1.19
1991	0.93	0.89–0.98
1992	1.04	1.00–1.09
1993	0.88	0.84–0.92
1994	1.14	1.08–1.21
1995	0.90	0.87–0.93
1996	1.06	1.02–1.10
1997	1.04	1.01–1.07
1998	0.94	0.91–0.97
1999	0.81	0.78–0.84
2000	1.12	1.08–1.17
2001	1.10	1.05–1.14

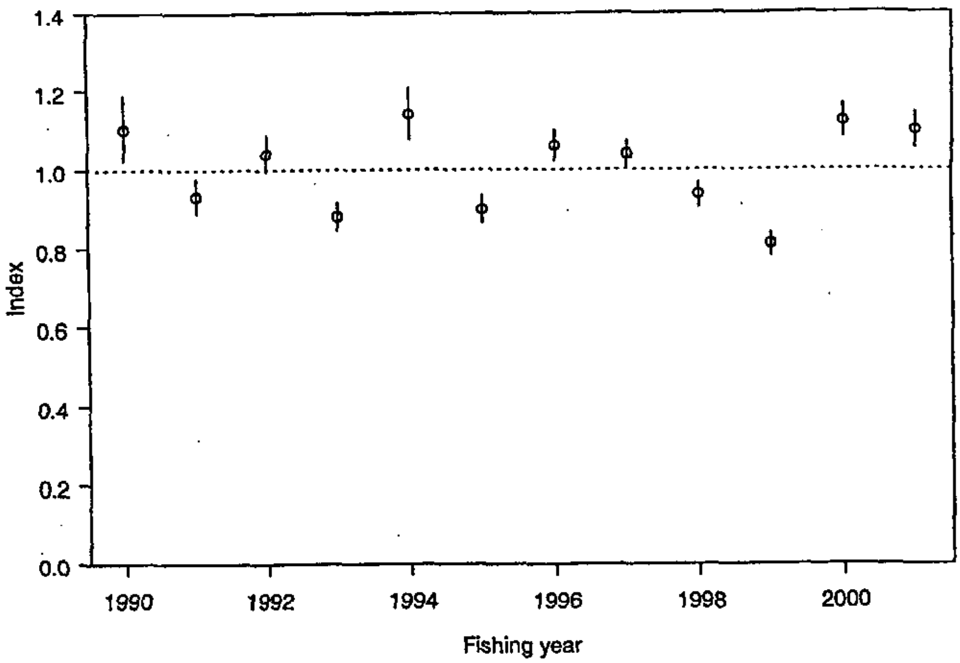

Figure 3: Relative year effects and 95% confidence intervals by fishing year using the mixed-effect model for the Chatham Rise fishery (includes Statistical Area 404).

The diagnostic plots for the mixed-effect model show some clustering in the residuals due to targeting and fishing in certain subareas (Appendix C, Figure C1, top two panels). From the normal quantile-quantile plots, the diagnostics for the fixed effects suggest that the model can be improved, and there may be violations of model assumptions (normality of residuals), but the diagnostics for the random effects appear to be adequate.

3.2 Chatham Rise stock (excluding Statistical Area 404)

This data set included 29 core vessels, catching 6297 t (37% of the total hake Chatham Rise catch) of hake; there were 24 707 records (65% of the total Chatham Rise records), (Table 5).

Table 5: Summary of the number of records, catch and number of unique vessels for core vessels fishing in the Chatham Rise fishery (excluding Statistical Area 404) by fishing year.

Fishing year	Number of records	Catch (t)	Number of vessels
1990	510	134	5
1991	1 885	501	11
1992	2 446	805	14
1993	1 886	388	15
1994	983	295	12
1995	1 732	360	12
1996	1 819	611	13
1997	2 642	563	16
1998	3 042	566	18
1999	3 214	728	17
2000	2 507	546	16
2001	2 041	800	14
Total	24 707	6 297	28

The variables selected by the stepwise regression for the lognormal and mixed-effect models are given in Table 6. Three variables were selected with a total r^2 of about 35% for the lognormal model, and five variables with a total r^2 of 31% for the mixed-effect model.

Table 6: Variables in order of selection by stepwise regression and the percentage of the deviance explained (r^2) for the lognormal and mixed-effect model for the Chatham Rise fishery (excludes Statistical Area 404).

Lognormal model		Mixed-effect model	
Selected variable	r^2	Selected variable	r^2
Vessel	22.0	*Subarea*	17.9
Subarea	34.3	*Target species*	25.1
Fishing year	35.4	*Fishing year*	27.5
		Net depth	28.7
		*Subarea*Net depth*	30.6

The resulting year effects using the lognormal and mixed-effect models are listed in Table 7 and plotted in Figure 4. The two sets of indices were similar and fluctuated from 1989–90 to 1995–96. They decreased for the next 3 years followed by an increase in 1999–00 and 2000–01.

Table 7: Relative year effects and 95% confidence intervals (CI) by fishing year using the lognormal and the mixed-effect models for the Chatham Rise fishery (excludes Statistical Area 404).

	Lognormal model		Mixed-effect model	
Fishing year	Index	95% CI	Index	95% CI
1990	0.98	0.89–1.08	0.98	0.92–1.05
1991	0.95	0.90–1.00	1.05	1.00–1.09
1992	1.14	1.08–1.21	1.13	1.08–1.17
1993	0.91	0.86–0.96	0.90	0.86–0.93
1994	1.09	1.02–1.16	1.02	0.97–1.07
1995	0.89	0.85–0.94	0.88	0.84–0.91
1996	1.04	0.99–1.10	1.04	1.00–1.08
1997	1.04	1.00–1.09	1.04	1.01–1.08
1998	0.95	0.91–0.98	0.96	0.93–0.99
1999	0.86	0.83–0.90	0.87	0.85–0.90
2000	1.07	1.02–1.12	1.07	1.03–1.10
2001	1.12	1.06–1.18	1.12	1.08–1.16

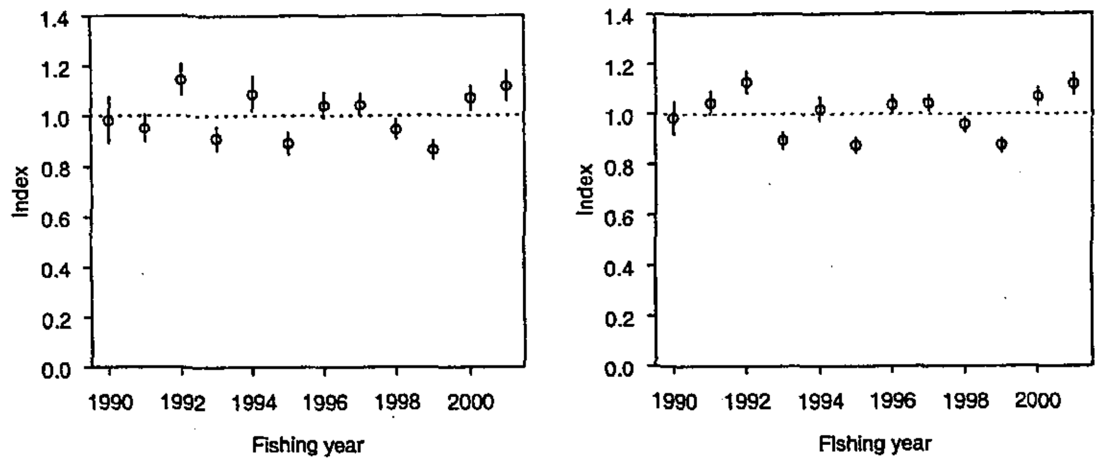

Figure 4: Hake CPUE indices and 95% confidence intervals by fishing year Chatham Rise fishery (excludes Statistical Area 404). Left, lognormal model, right, mixed-effect model.

The effects of the selected variables for the lognormal model are shown in Figure 5. Twenty-eight vessels were incorporated in the model, but one had a very high catch rate. The difference between the highest and lowest catching vessels was a factor of about 5. Highest catch rates tend to occur in the subareas on the north side of the Chatham Rise.

The diagnostic plots for the different models and data sets are presented in Appendix D, Figures D1–D3. The diagnostics for the fixed effects indicate a reasonable pattern in the residuals (top panels Figure D1), but the quantile-quantile plots indicate a deviation from the normal distribution of the residuals at either end, suggesting that low and high catch rates were not modelled well. This suggests that the models can be improved, and there may be violations of model assumptions. The quantile-quantile plots for the random effects of the mixed-effect model appear to be adequate.

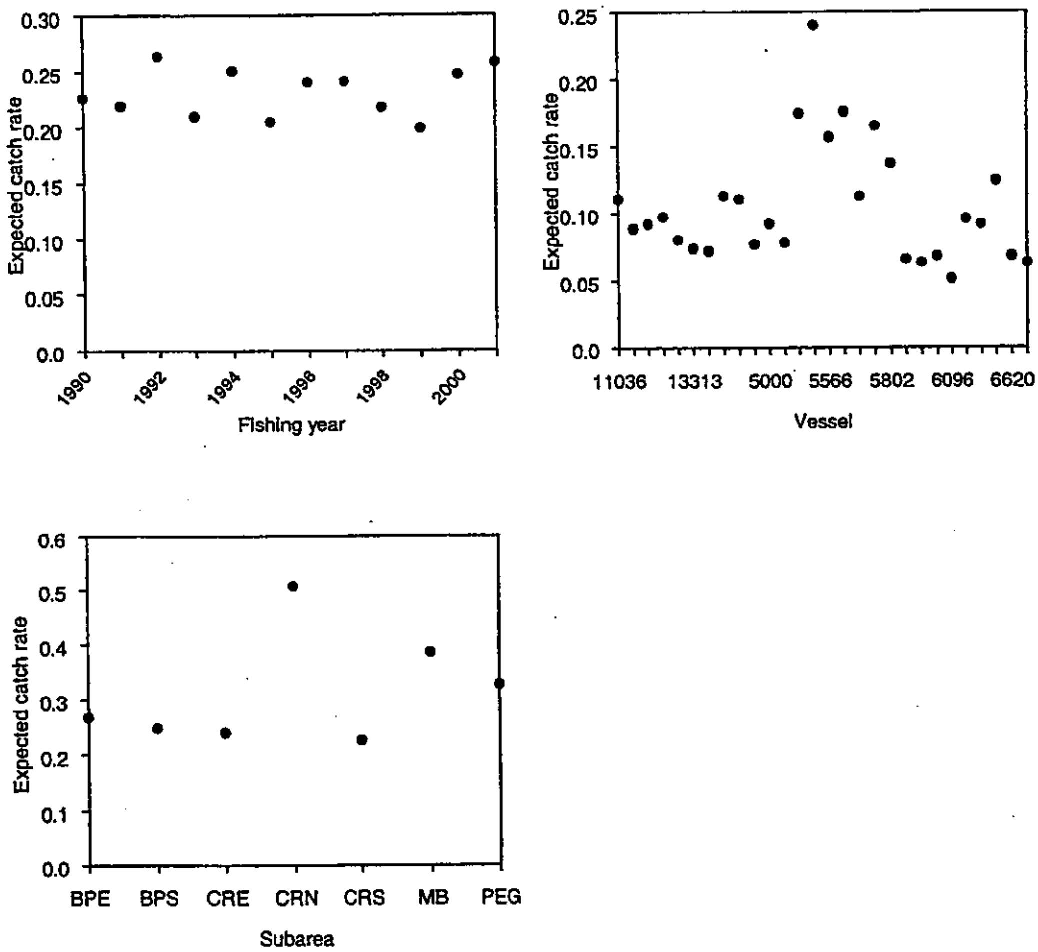

Figure 5: Expected catch rates (t per tow) for median values of fixed parameters using the lognormal model for the Chatham Rise fishery (excludes Statistical Area 404). Note, model used positive tows only.

3.3 Statistical Area 404 fishery

By restricting the data set to tows from Statistical Area 404, 64 vessels were included, which resulted in 2214 records (6% of the total records), and 8811 t of catch (52% of the total catch). By further restricting the data to core vessels, eight unique vessels were included, which resulted in 1230 records (3% of the total records) and 5965 t of catch (35% of the total catch) (Table 8).

Table 8: Summary of the number of records, catch and number of unique vessels for core and all vessels fishing in the Statistical Area 404 fishery by fishing year.

	All vessels			Core vessels		
Fishing year	No. records	Catch (t)	No. vessels	No. records	Catch (t)	No. vessels
1990	89	429	5	-	-	-
1991	44	93	4	-	-	-
1992	140	503	10	75	285	2
1993	270	644	11	79	254	3
1994	124	708	12	22	68	3
1995	372	1 233	21	147	720	5
1996	277	1 774	12	198	1 532	5
1997	220	1 078	12	184	1 006	5
1998	136	455	16	89	364	8
1999	194	677	13	177	653	7
2000	81	217	10	65	209	4
2001	267	1 000	17	194	874	5
Total	2 214	8 811	64	1 230	5 965	8

The variables selected by the stepwise regression for the lognormal and mixed-effect models are given in Table 9. Six variables were selected with a total r^2 of about 24% for the lognormal model, and 6 variables with a total r^2 of 25% for the mixed-effect model. The variables selected were similar (*vessel* is a random effect in the mixed-effect model, therefore it cannot be selected).

Table 9: Variables selected by order of selection by stepwise regression and the percentage of the deviance explained (r^2) for the lognormal and mixed-effect models for the Statistical Area 404 fishery.

Lognormal model		Mixed-effect model	
Selected variable	r^2	Selected variable	r^2
Fishing year	6.3	*Fishing year*	6.3
Distance	12.3	*Distance*	11.2
Month	15.7	*Month*	15.6
Vessel	19.2	*Start time*	18.9
Start time	21.7	*Start longitude*	22.3
Start longitude	24.1	*Month*Distance*	24.5

The resulting year effects from the different models are listed in Table 10 and are plotted in Figure 6. The three sets of indices were similar. They showed a declining trend, with increases in 1994–95, 1995–96, and 2000–01.

Table 10: Relative year effects and 95% confidence intervals (CI) by fishing year using the lognormal and mixed-effect models for the Statistical Area 404 fishery.

	Lognormal model		Mixed-effect model	
Fishing year	Index	95% CI	Index	95% CI
1992	1.77	1.26–2.49	1.50	1.05–2.15
1993	1.21	0.94–1.57	1.10	0.85–1.42
1994	1.00	0.66–1.53	0.93	0.61–1.42
1995	1.25	1.03–1.52	1.33	1.09–1.64
1996	1.40	1.18–1.66	1.53	1.29–1.83
1997	0.89	0.74–1.06	0.90	0.75–1.08
1998	0.61	0.50–0.76	0.68	0.54–0.84
1999	0.72	0.60–0.87	0.75	0.63–0.90
2000	0.53	0.41–0.69	0.57	0.44–0.74
2001	1.27	1.05–1.55	1.23	1.01–1.50

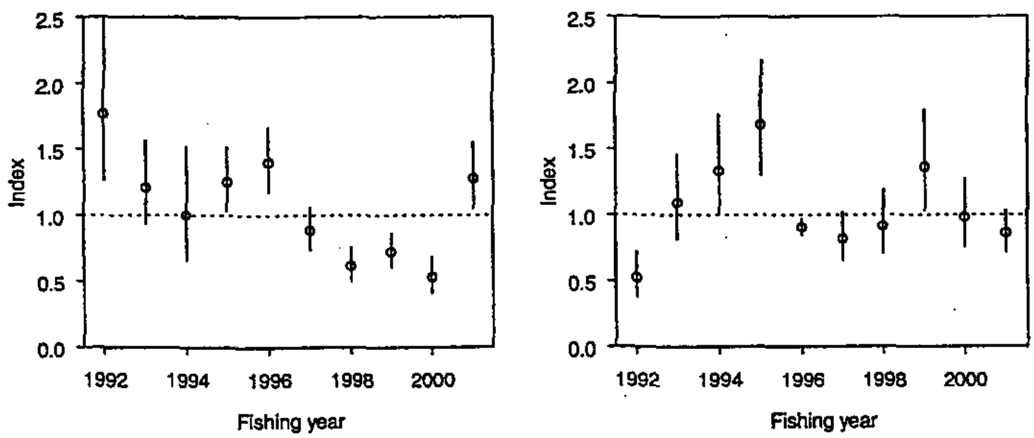

Figure 6: Hake CPUE indices and 95% confidence intervals by fishing year for the Statistical Area 404 fishery from the lognormal model (left), and the mixed-effect model (right).

The effects of the selected variables are shown in Figure 7. The difference between the least and most efficient hake catching vessels was a factor of about 2. Highest catch rates tended to occur where tows were less than 5 n. miles long. Catch rates were also higher from tows that were deployed midmorning, and those in mid longitude of Statistical Area 404.

The diagnostic plots for the different models (Appendix E, Figures E1–E2) indicate a reasonable pattern in the residuals for both the mixed-effect and lognormal models, but the quantile-quantile plots indicate a deviation from the normal distribution of the residuals at the lower end, suggesting that very low values of catch rate were not well predicted. This suggests the models can be improved, and there may be violations of model assumptions. The diagnostics for the random effects of the mixed-effect model appeared adequate.

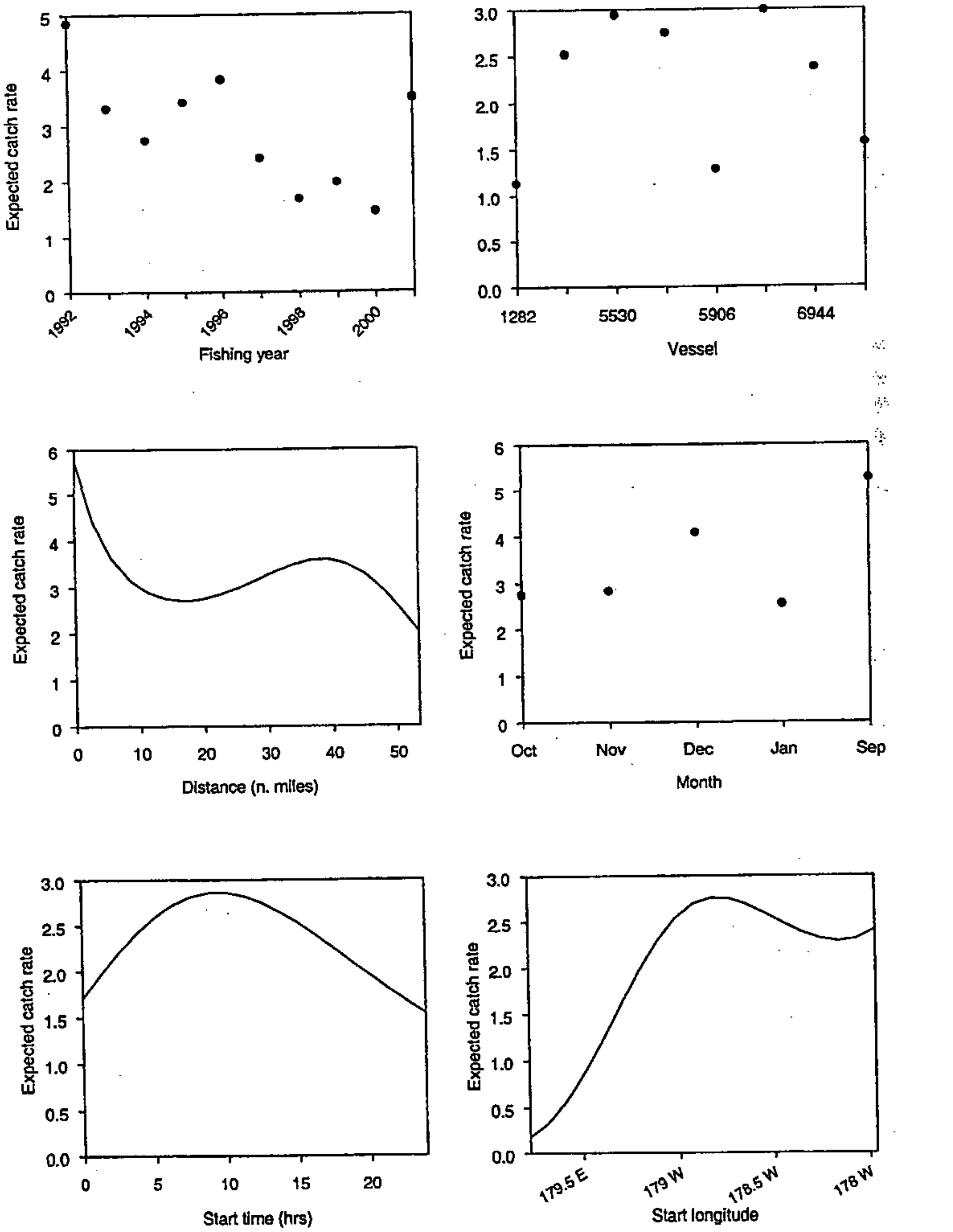

Figure 7: Expected catch rates (t per tow) for median values of fixed parameters using the lognormal model for core vessels fishing in the Statistical Area 404 fishery on the Chatham Rise.

3.4 Sub-Antarctic fishery

The Sub-Antarctic fishery was fished by 151 unique vessels and caught 20 683 t hake. The data set comprised 33 058 records (Table 11). Restricting the data to core vessels resulted in a 42% reduction in records to 19 018 records and a 21% reduction in hake catch to 16 321 t (15 vessels). By further restricting the data to the Stewart/Snares spawning area, there was a reduction to 2009 records, 8827 t of hake catch (6 vessels).

Table 11: Summary of the number of records, hake catch, and number of unique vessels for core vessels and all vessels fishing in the Sub-Antarctic fishery.

	Sub-Antarctic all vessels			Sub-Antarctic core vessels			Stewart/Snares core vessels		
Fishing year	No. Records	Catch	No. Vessels	No. Records	Catch	No. Vessels	No. Records	Catch	No. Vessels
1990	1 372	1 410	32	547	1 094	5	38	65	2
1991	2 673	1 436	31	1 262	916	7	20	7	3
1992	4 018	1 942	41	2 756	1 561	9	12	5	3
1993	3 801	2 374	46	2 737	1 632	9	149	429	3
1994	2 004	753	42	952	326	9	117	147	5
1995	2 107	1 340	38	1 237	1 165	7	140	949	4
1996	1 635	2 199	34	738	2 015	7	185	1 803	4
1997	2 434	1 551	50	1 338	1 294	10	171	958	5
1998	3 017	1 671	46	2 079	1 509	9	263	1 008	4
1999	2 433	1 767	52	1 471	1 514	8	287	1 221	4
2000	3 819	2 247	51	2 010	1 671	9	308	1 077	4
2001	3 745	1 994	53	1 891	1 623	9	319	1 157	4
Total	33 058	20 683	151	19 018	16 321	15	2 009	8 827	6

The variables selected by the stepwise regression for the analysis of the Sub-Antarctic and Stewart/Snares box fisheries using the core vessels for the lognormal and mixed-effect models are given in Table 12. For the Sub-Antarctic fishery, 6 variables were selected with a total r^2 of about 54% using the lognormal model, and 5 variables were selected with a total r^2 of 52% using the mixed-effect model. Similar variables were selected (vessel was not offered in the mixed-effect model as this was a random factor). For the Stewart/Snares box fishery, 5 variables were selected with a total r^2 of 60%, and again similar variables were selected.

Table 12: Variables selected by order of selection and the percentage of the deviance explained (r^2) using stepwise regression for the lognormal and mixed-effect models for the Sub-Antarctic and Stewart/Snares fisheries, core vessels, (*, interaction terms).

Sub-Antarctic				Stewart/Snares	
Lognormal model		Mixed-effect model		Mixed-effect model	
Selected variable	r^2	Selected variable	r^2	Selected variable	r^2
Vessel	31.2	*Target species*	28.5	*Month*	41.0
Month	38.1	*Month*	36.7	*Target species*	51.5
Subarea	42.9	*Subarea*	42.4	*Start longitude*	56.0
*Subarea*Month*	50.5	*Subarea*Month*	49.3	*Target species*	58.4
Target species	52.7	*Fishing year*	51.6	*Target species*Month*	59.6
Fishing year	53.8				

The resulting year effects using the different models are listed in Table 13 and are plotted in Figure 8. The three sets of indices from the Sub-Antarctic fishery were similar with a decline up to 1994–95, increasing in 1995–96, declining in 1996–96, and no trend from 1996–97 on. The indices from the Stewart/Snares area fishery showed a similar trend to the Sub-Antarctic series, but the decline from 1993–94 to 1994–95 was much steeper.

Table 13: Relative year effects and 95% confidence intervals by fishing year using the lognormal and mixed-effect models for the Sub-Antarctic and Stewart/Snares fisheries.

	Sub-Antarctic				Stewart/Snares	
	Lognormal model		Mixed-effect model		Mixed-effect model	
Fishing year	Index	95% CI	Index	95% CI	Index	95% CI
1990	1.45	1.32–1.60	1.46	1.34–1.58	-	-
1991	1.31	1.22–1.40	1.31	1.23–1.38	-	-
1992	1.22	1.16–1.29	1.22	1.17–1.27	-	-
1993	1.03	0.97–1.08	1.18	1.13–1.23	3.46	2.54–4.70
1994	0.90	0.84–0.97	0.84	0.79–0.90	1.23	0.91–1.67
1995	0.83	0.78–0.88	0.83	0.79–0.87	1.04	0.85–1.26
1996	1.29	1.19–1.39	1.11	1.05–1.19	1.77	1.48–2.12
1997	0.91	0.86–0.96	0.83	0.79–0.87	0.92	0.76–1.12
1998	0.89	0.84–0.95	0.89	0.85–0.93	0.82	0.69–0.98
1999	0.78	0.73–0.83	0.79	0.75–0.83	0.67	0.57–0.80
2000	0.85	0.80–0.91	0.94	0.89–0.99	0.49	0.42–0.59
2001	0.81	0.76–0.86	0.87	0.82–0.91	0.50	0.43–0.60

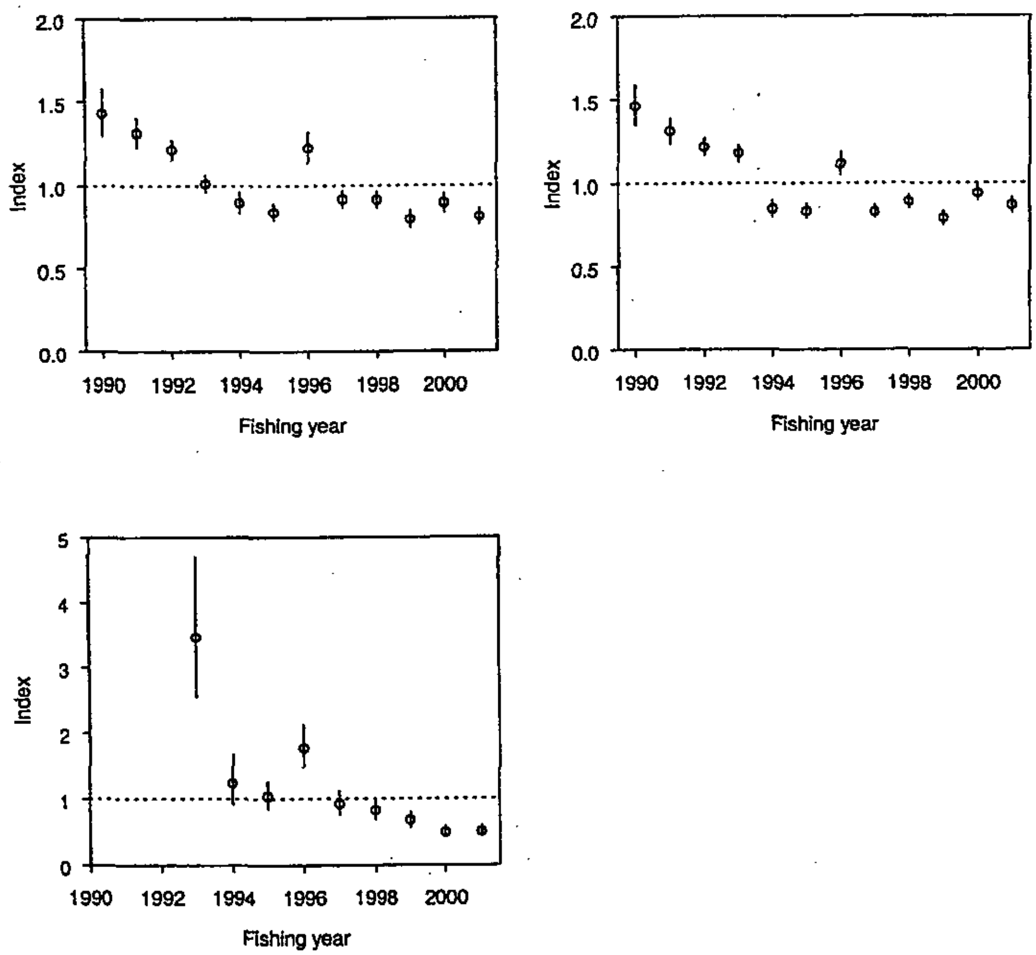

Figure 8: Hake CPUE indices and 95% confidence intervals by fishing year for the Sub-Antarctic fishery. Top left, lognormal model; top right, mixed-effect model; bottom left, mixed-effect model using data from Stewart/Snares fishery only.

The effects of the selected variables are shown in Figure 9. The difference between the least and most efficient catching vessels was a factor of about 5. Highest catch rates tended to occur in August/September and in the north and south *subareas* while targeting hake.

The diagnostic plots for the different models and data sets are presented in Appendix F, Figures F1–F3. The diagnostics for both the lognormal model and fixed effects of the mixed-effect model indicated a reasonable pattern in the residuals, but the quantile-quantile plots indicated a deviation from the normal distribution of the residuals at either end, suggesting that very small and large values of catch rate were not well modelled. This suggests that the models can be improved, and there may be violations of model assumptions. The diagnostics for the random effects of the mixed-effect models appear to be adequate, apart from the Stewart/Snares fishery, where deviations occurred at either end.

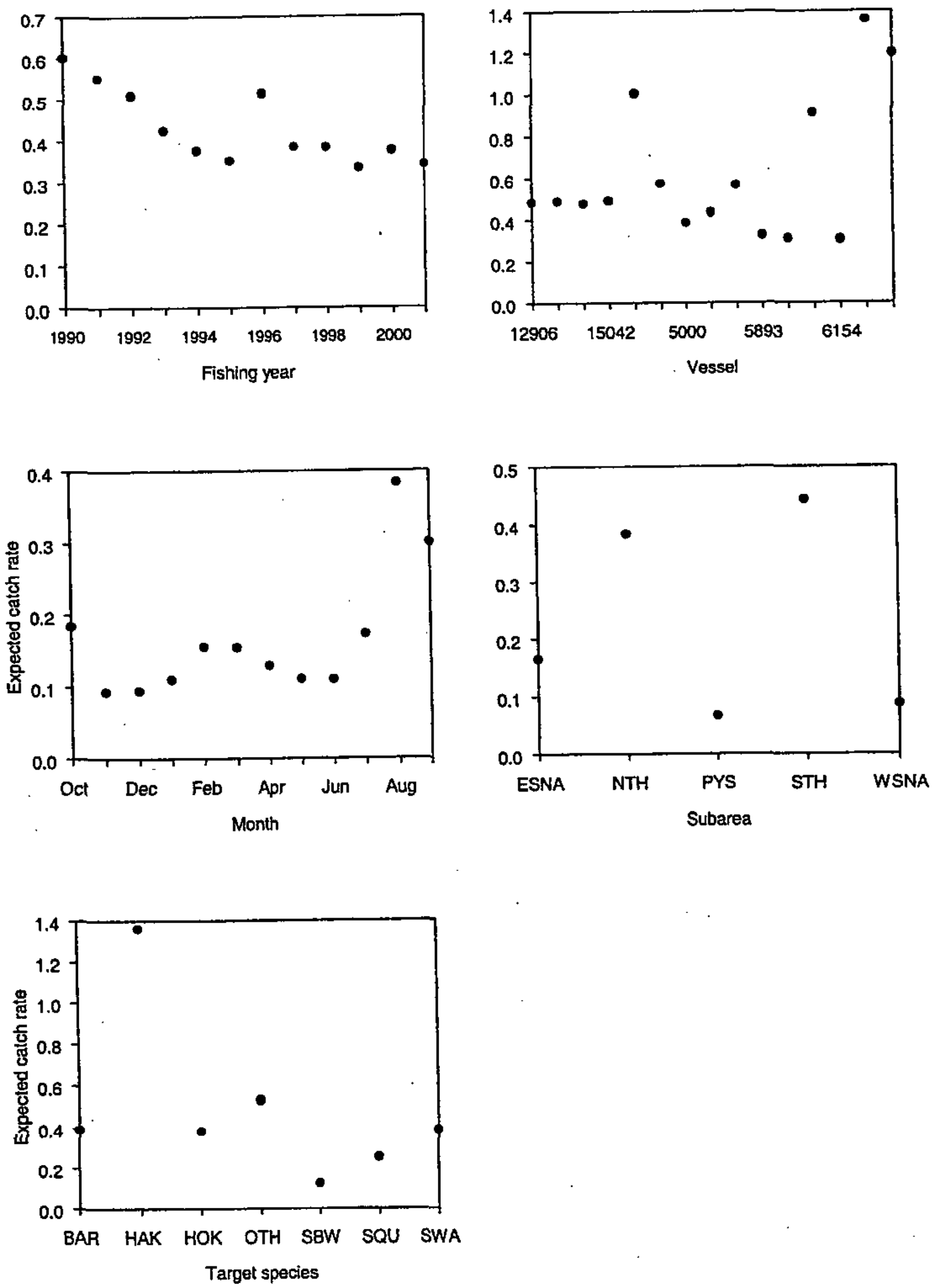

Figure 9: Expected catch rates (t per tow) for median values of fixed parameters using the lognormal model for the Sub-Antarctic fishery. Definitions of sub-areas can be found in Appendix A, Figure A2

4. DISCUSSION

In a review of the calculation and interpretation of CPUE indices in general, Dunn et al. (2000a) noted that calculation of CPUE indices does not necessarily result in an index which is related to abundance. They cautioned against the use of CPUE indices in stock assessment models until several aspects of the analysis had been evaluated and the CPUE indices themselves had been validated by fishery-independent data. They recommended that CPUE analyses should include a discussion of 1) a definition of the relationship between CPUE and fish abundance, 2) an assessment of data adequacy, 3) the methods of model fitting and validation, and 4) evaluation of the CPUE index in an attempt to validate the data selection, model method, and results. These points are discussed below with respect to the hake CPUE analyses presented here.

4.1 The relationship between CPUE and hake abundance

For the analysis and interpretation of the indices we have assumed a simple direct relationship between CPUE and abundance. There were, however, specific areas and times (e.g., Statistical Area 404 during the spawning season) where hake were more abundant and hence targeted. There appears to be little significant searching carried out by vessels or the targeting of marks, both of which could lead to a hyperstable CPUE/abundance relationship (Dunn et al. 2000a). Mean catch rates for hake were 468 kg per tow in the Chatham Rise fishery, 3.8 t per tow in Statistical Area 404 fishery, and 670 kg per tow in the Sub-Antarctic fishery, which are all low enough to suggest that trawl saturation is not usually a problem. It is also unlikely that catch per unit effort was artificially high due to targeting a spawning population, and the variable *month* selected in some of the models was likely to correct for any differences in catch rates between spawning and non-spawning seasons.

4.2 Assessment of data adequacy

The data used in the analysis were carefully groomed to remove tows in which catch was misreported, could have been misreported, and other errors such as incorrect values. Although some errors may still exist, they are likely to have a negligible effect on the CPUE analysis.

4.3 Model fitting and validation

Model fitting and model validation were considered by comparing the explanatory variables, the variation explained, the diagnostic plots, and the results of the different models used in the analysis.

Vessel or *subarea* appeared to be important predictors as they were selected in the lognormal models, and *subarea* and *target species* were selected for the mixed-effect models. The resulting residual variance explained ranged from 24% for the Statistical Area 404 fishery to 33% for the Chatham Rise fishery and about 51% for the Sub-Antarctic mixed fishery, with the first two or three variables providing most of the explanatory power.

For the Chatham Rise and Statistical Area 404 fisheries, the variables *subarea* and *target species* entered the model at an early stage. This corresponds with our understanding of the fishery, where vessels targeted hake in specific areas. However *month* or *fishing day* was not selected, but Phillips (unpublished results) noted that 45–93% of the hake catch in the

Chatham Rise fishery was taken from September to January, which corresponds to the hake spawning period. It is possible that *month* is confounded with *target species*. *Month* entered the model from the Statistical Area 404 and the Sub-Antarctic fisheries, which corresponds with the spawning period of hake.

The CPUE indices from Statistical Area 404 and the wider Chatham Rise fisheries differ from each other (Figure 6). All indices from this study differ from those reported by Kendrick (unpublished analysis) as ours showed declines in the 1995–96 and 1996–97 fishing years. These differences may have been because Kendrick (unpublished analysis) only used hake targeted hake tows in her analysis.

The generalised linear modelling approach of Gavaris (1980) allows standardisation of a range of factors, but only those that can be derived from TCEPR returns. Other factors that may affect fishing, for example, changes in gear, improved technology, skipper's skills, and local knowledge, were not considered (Dunn 2002).

The use of mixed-effect models has proved effective in calculating the CPUE indices, but a method needs to be developed so records with zero catches can be included in the analysis.

4.4 Evaluation of the CPUE index

An important step in assessing the usefulness of CPUE as an index of abundance is to determine whether annual changes in CPUE reflect the abundance of hake in the fisheries.

A time series of annual trawl surveys initiated on the Chatham Rise in 1992 (Stevens et al. 2001) overlaps with 10 of the fishing years analysed here. Trawl survey hake abundance indices show a decline within the time series, and although this correlates with the declining catch, it does not correlate well with the CPUE indices (which show no trend) derived from the Chatham Rise fishery (Appendix G, Figure G1). CPUE indices derived from Statistical Area 404, however, show a reasonable correlation with the trawl survey abundance indices (extracted from the relevant strata), apart from 1996 (Figure G2). However, there is some concern as the rates of change in the Statistical Area 404 CPUE indices showed a large decline from 1996 to 1997 and a large increase in 2000 to 2001. Such large fluctuations in CPUE may render them unreliable as indices of abundance and caution is advised.

Two series of trawl surveys have assessed hake abundance in the Sub-Antarctic; autumn trawl surveys in 1992, 1993, 1995, and 1998, and summer trawl surveys in 1991–1993, 2000, and 2001. Both these series show some correlation with the CPUE indices, apart from in 1992 and 1996 (Figure G3), and may therefore reflect hake abundance.

Previous and current CPUE indices were similar and were unable to capture the extremes in catch rate observed in the fishery. Predictive models generally underestimated the observed catches at the upper extreme of the observed range of catches, while overestimating the catch from trawls at the lower range. Such models are reported to have poor fits, and model diagnostics have suggested serious departure from model assumptions. Grouping of the fitted values is also apparent for the Chatham Rise and Sub-Antarctic fisheries, due to the different catch rates for each target species and subarea.

The use of mixed-effect models in this analysis reduced the number of parameters selected. The resulting indices were very similar to the indices derived from conventional methods. Further work could, however, be done. For example, the fixed variables were selected using a forward stepwise multiple regression-fitting algorithm, which may not be the best method for selecting the fixed effects. Other work should consider a method to include zero tows.

This report is part of an on-going project to develop CPUE indices for estimating hake abundance for the Chatham Rise and Sub-Antarctic hake fisheries. Different methods were presented, using different data sets. However, more work needs to be done improving model diagnostics and residuals. This could involve using data sets based on vessel processing type, as used in the CPUE analysis of the west coast hoki fishery (Langley et al. 2001). The exclusion of records with zero catch may also have an effect on the resulting indices.

5. CONCLUSIONS

- CPUE indices from the Statistical Area 404 mixed fishery are consistent with trawl survey indices of hake, but there are large increases and decreases between some years, and therefore caution is advised.
- CPUE indices from the Chatham Rise fishery are inconsistent with trawl survey abundance estimates of hake and may not reflect abundance.
- CPUE indices from the Sub-Antarctic fishery show some correlation with the abundance indices from Sub-Antarctic trawl surveys, apart from the 1992 and 1996 fishing year, and may reflect abundance.
- The use of mixed-effect models has potential advantages over the more traditionally applied linear models, as fewer variables need to be selected. This can prevent model overfitting and CPUE indices are less prone to the bias caused by unbalanced data sets, but further improvements (recommended above) could be made.

6. ACKNOWLEDGMENTS

This work was funded by the Ministry of Fisheries under Project HAK 2001/01 (Objective 1). Thanks to Murray Smith for help with mixed-effect models, and Owen Anderson for reviewing this manuscript.

7. REFERENCES

Annala, H.; Sullivan, K.J.; O'Brien, C.J.; Smith, N.M.W.M.; Grayling, S.M. (comp.) (2003). Report from the Fishery Assessment Plenary, May 2003: stock assessments and yield estimates 616 p. (Unpublished report held in NIWA library, Wellington.)

Barrowman, N.J. (2000). Nonlinear Mixed Effects Models for Meta-Analysis. Unpublished Ph.D. thesis. Dalhousie University, Halifax, Nova Scotia, Canada. 175 p.

Chambers, J.M.; Hastie, T.J. (1991). Statistical models in S. Wadsworth and Brooks/Cole, Pacific Grove, CA. 608 p.

Colman, J.A. (1988) Hake. New Zealand Fisheries Assessment Research Document 88/31. 23 p. (Unpublished report held in NIWA library, Wellington.)

Colman, J.A. (1995). Biology and fisheries of New Zealand hake (*Merluccius australis*). *In*: Hake: biology, fisheries and markets. Alheit, J.; Pitcher, T. J. (eds). *(Fish and Fisheries Series 15)* pp 365–388. Chapman and Hall, London.

Colman, J.A.; Stocker, M.; Pikitch, E. (1991) Assessment of hake (*Merluccius australis*) stocks for the 1991–92 fishing year. New Zealand. Fisheries Assessment Research Document 91/14. 29 p. (Unpublished report held in NIWA library, Wellington.)

Dunn, A. (2002). Updated catch-per-unit-effort indices for hoki (*Macruronus novaezelandiae*) on the west coast South Island, Cook Strait, Chatham Rise, and sub-

Antarctic for the years 1990 to 2001. *New Zealand Fisheries Assessment Report 2002/47* 51 p.

Dunn, A. (2003). Revised estimates of landings of hake (*Merluccius australis*) for the west coast South Island, Chatham Rise, and sub-Antarctic stocks in the fishing years 1989–90 to 2000–01. *New Zealand Fisheries Assessment Report 2003/36* 36 p.

Dunn, A.; Harley, S.J.; Doonan, I.J.; Bull, B. (2000a). Calculation and interpretation of catch-per-unit-effort (CPUE) indices. *New Zealand Fisheries Assessment Report 2000a/1.* 44 p.

Dunn, A.; Horn, P.L.; Cordue, P.L.; Kendrick, T.H. (2000b). Stock assessment of hake (*Merluccius australis)* for the 1999–2000 fishing year. *New Zealand Fisheries Assessment Report 2000/50.* 50 p.

Francis, R.I.C.C. (2001). Orange roughy CPUE on the South and East Chatham Rise. *New Zealand Fisheries Assessment Report 2001/26.* 30 p.

Gavaris, S. (1980). Use of a multiplicative model to estimate catch rate and effort for commercial data. *Canadian Journal of Fisheries and Aquatic Sciences 37*: 2272–2275.

Kendrick, T.H. (1998) Feasibility of using CPUE as an index of stock abundance for hake. New Zealand Fisheries Assessment Research Document 98/27. 22 p. (Unpublished report held in NIWA library, Wellington.)

Langley, A.D.; Dunn, A.; Ballara, S.L. (2001). Determination of CPUE indices from catch effort data for hoki (*Macruronus novaezelandiae*) for the 1986–87 for 1998–99 fishing years. *New Zealand Fisheries Assessment Report 2001/49.* 70 p.

Pinheiro, J.C.; Bates, D.M. (2000). Mixed effects models in S and S-Plus. Springer, New York. 552 p.

Richardson, A.M.; Welsh, A.H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics 51*: 1429–1439.

Searle, S.R. (1987). Linear models for unbalanced data. Wiley series in probability and mathematical statistics. Applied probability and statistics. 536 p.

Stevens, D.W.; Livingston, M.E.; Bagley, N.W. (2001). Trawl survey of hoki and middle depth species on the Chatham Rise, January 2000 (TAN0001). *NIWA Technical Report 104.* 55 p.

Venables, W.N.; Ripley, B.D. (1994). Modern applied statistics with S-PLUS. 2nd ed. Springer-Verlag, New York. 462 p.

Vignaux, M. (1992) Catch per unit of effort (CPUE) analysis of the hoki fishery. New Zealand Fisheries Assessment Research Document 92/14. 30 p. (Unpublished report held in NIWA library, Wellington.)

Vignaux, M. (1994) Catch per unit of effort (CPUE) analysis of west coast South Island and Cook Strait spawning hoki fishery, 1987–93. New Zealand Fisheries Assessment Research Document 94/11. 29 p. (Unpublished report held in NIWA library, Wellington.)

Watson, G.S. (1966). Smooth regression analysis. *Sankha, Ser. A 26*: 359–378.

Appendix A: Definition of subareas

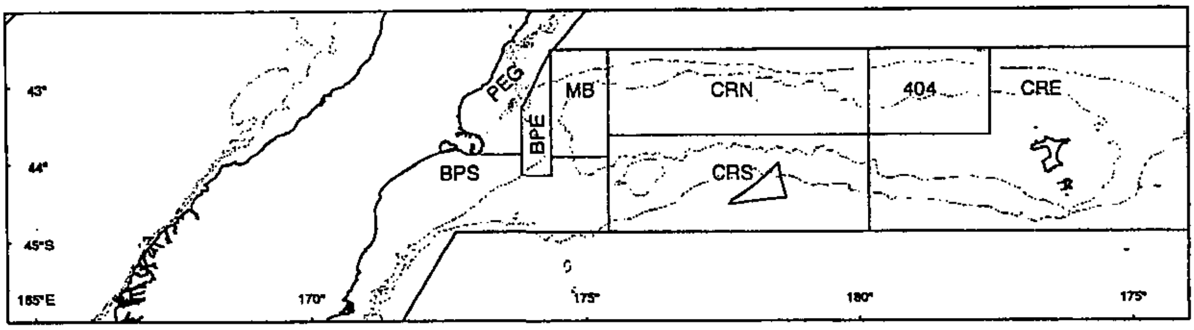

Figure A1: Area boundaries, and subareas for all tows in the Chatham Rise fishery. "404" is Statistical Area 404, "CRE" is Chatham Rise east, "CRN" is Chatham Rise north, "CRS" is Chatham Rise south, "MB" is Mernoo Bank, "BPE" is Banks Peninsula east, "BPS is Banks Peninsula south and "PEG" is Pegasus bay.

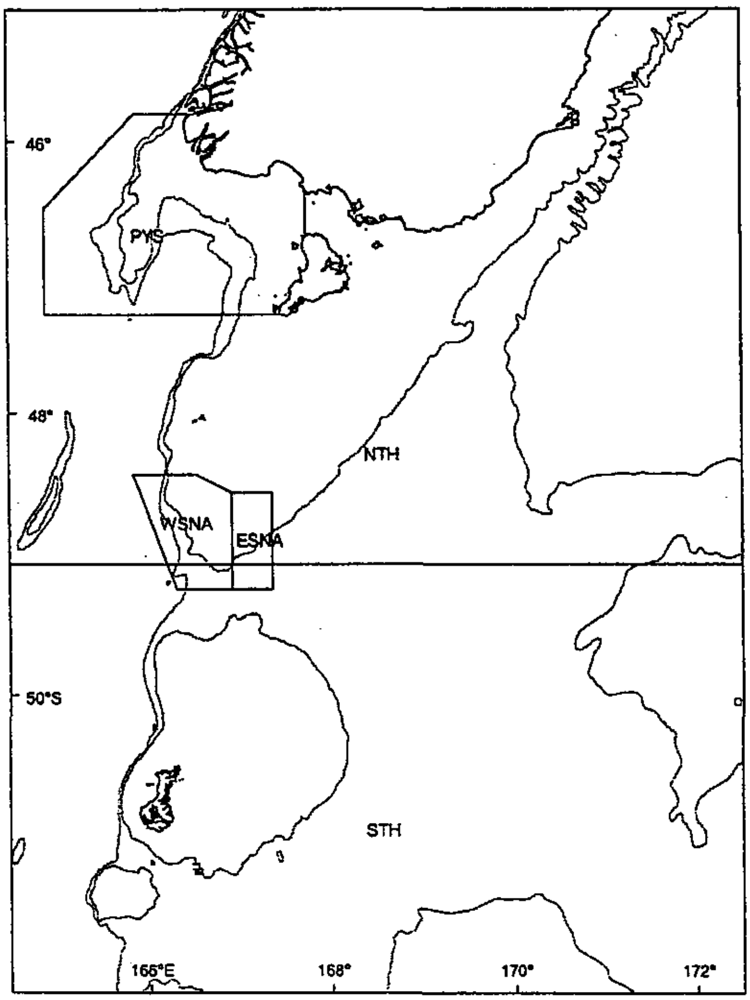

Figure A2: Area boundaries, and subareas for all tows in the Sub-Antarctic fishery. "PYS" is Pysaguer, "WSNA" is the western Snares, "ESNA" is the eastern Snares, "NHOLE" is the Norwegian Hole, "NTH" is all other areas north of latitude -46.36, "STH" is all other areas soulth of -46.36.

Appendix B: Vessel summary

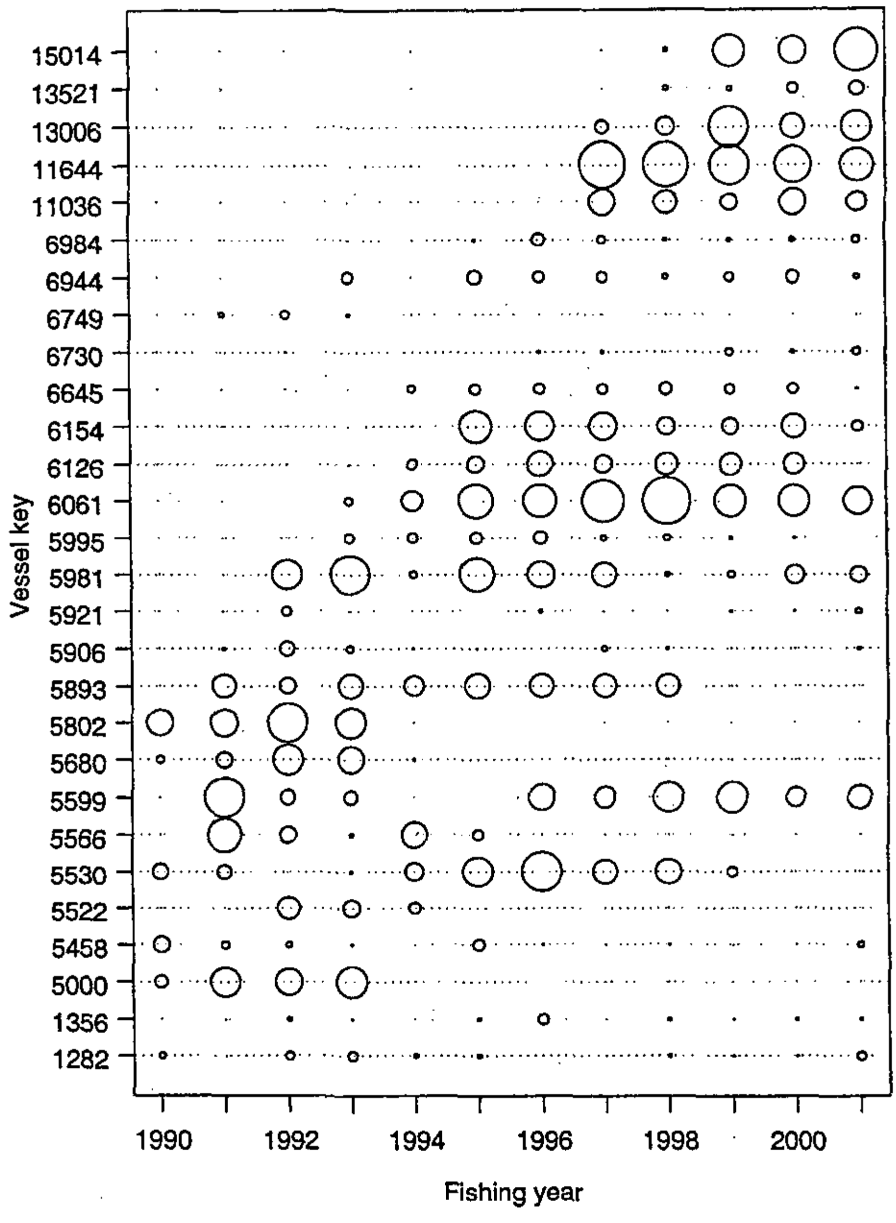

Figure B1: Summary of the number of tows by year (symbol area proportional to the number of tows) for the Chatham Rise fishery (includes Statistical Area 404), core vessels. Vessel key is a unique vessel identifying number. For the total number of tows by fishing year, refer to Table 2.

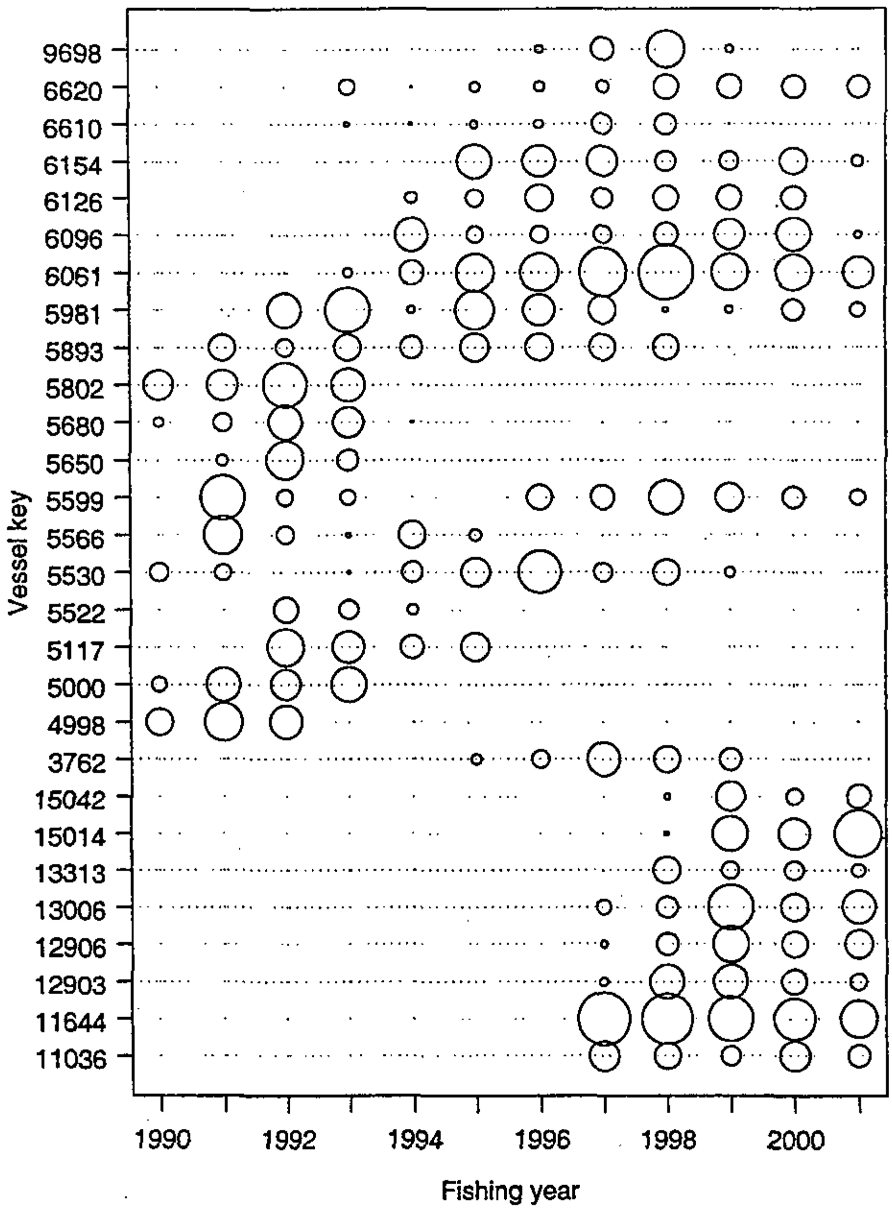

Figure B2: Summary of the number of tows by year (symbol area proportional to the number of tows) for the Chatham Rise fishery (excludes Statistical Area 404), core vessels. Vessel key is a unique vessel identifying number. For the total number of tows by fishing year, refer to Table 5.

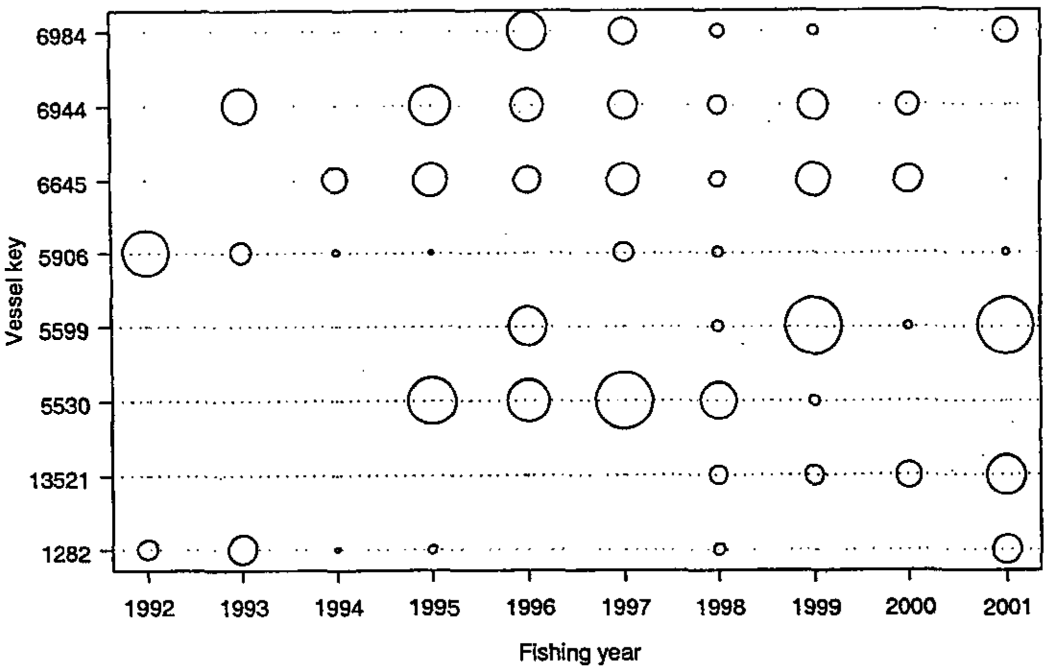

Figure B3: Summary of the number of tows by year (symbol area proportional to the number of tows) for core vessels involved in the Statistical Area 404 fishery. Vessel key is a unique vessel identifying number. For the total number of tows by fishing year, refer to Table 8.

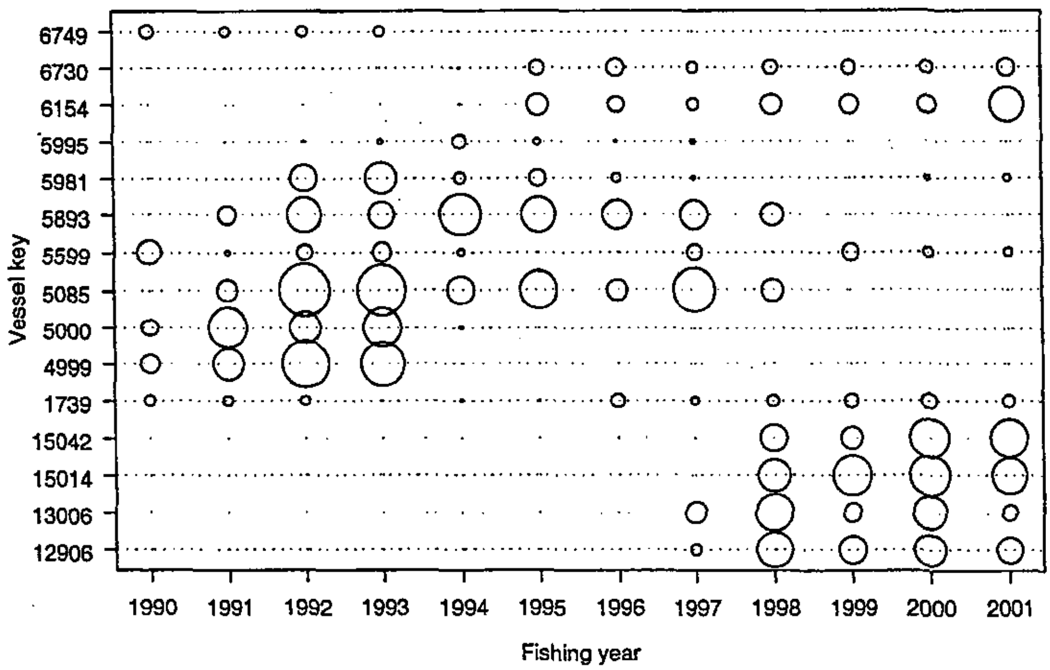

Figure B4: Summary of the number of tows by year (symbol area proportional to the number of tows) for core vessels involved in the Sub-Antarctic. For the total number of tows by fishing year, refer to Table 11.

Appendix C: Diagnostic plots, Chatham Rise fishery (includes Statistical Area 404)

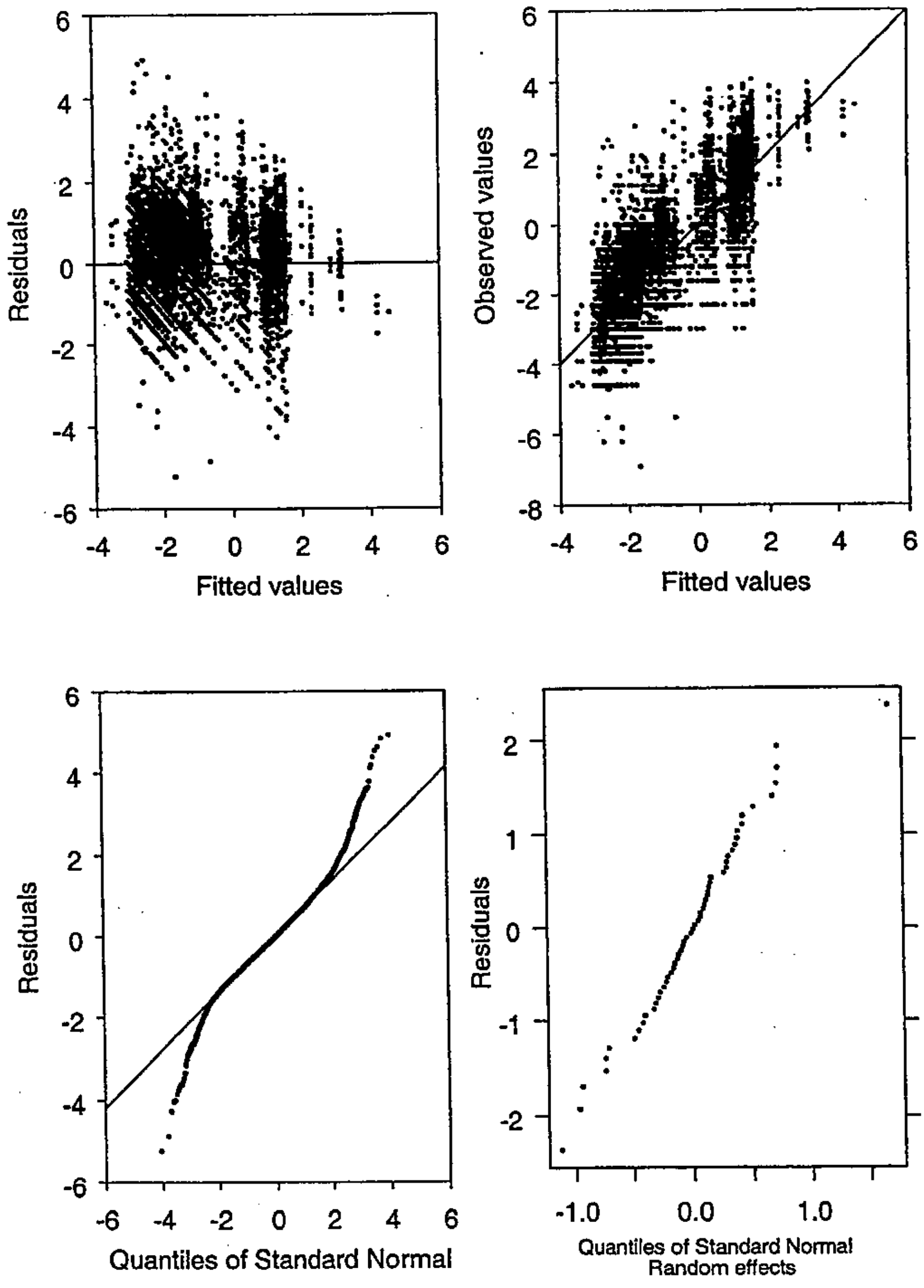

Figure C1: Diagnostic plots (log scale) of the fixed effects of mixed-effect model for the Chatham Rise fishery (including Statistical Area 404). Top left, fitted values versus residuals; top right, fitted values versus observed values; bottom left, normal quantile-quantile plot of the residuals; bottom right, a normal quartile-quartile plot of the residuals for method nested in vessel.

Appendix D: Diagnostic plots, Chatham Rise fishery (excludes Statistical Area 404)

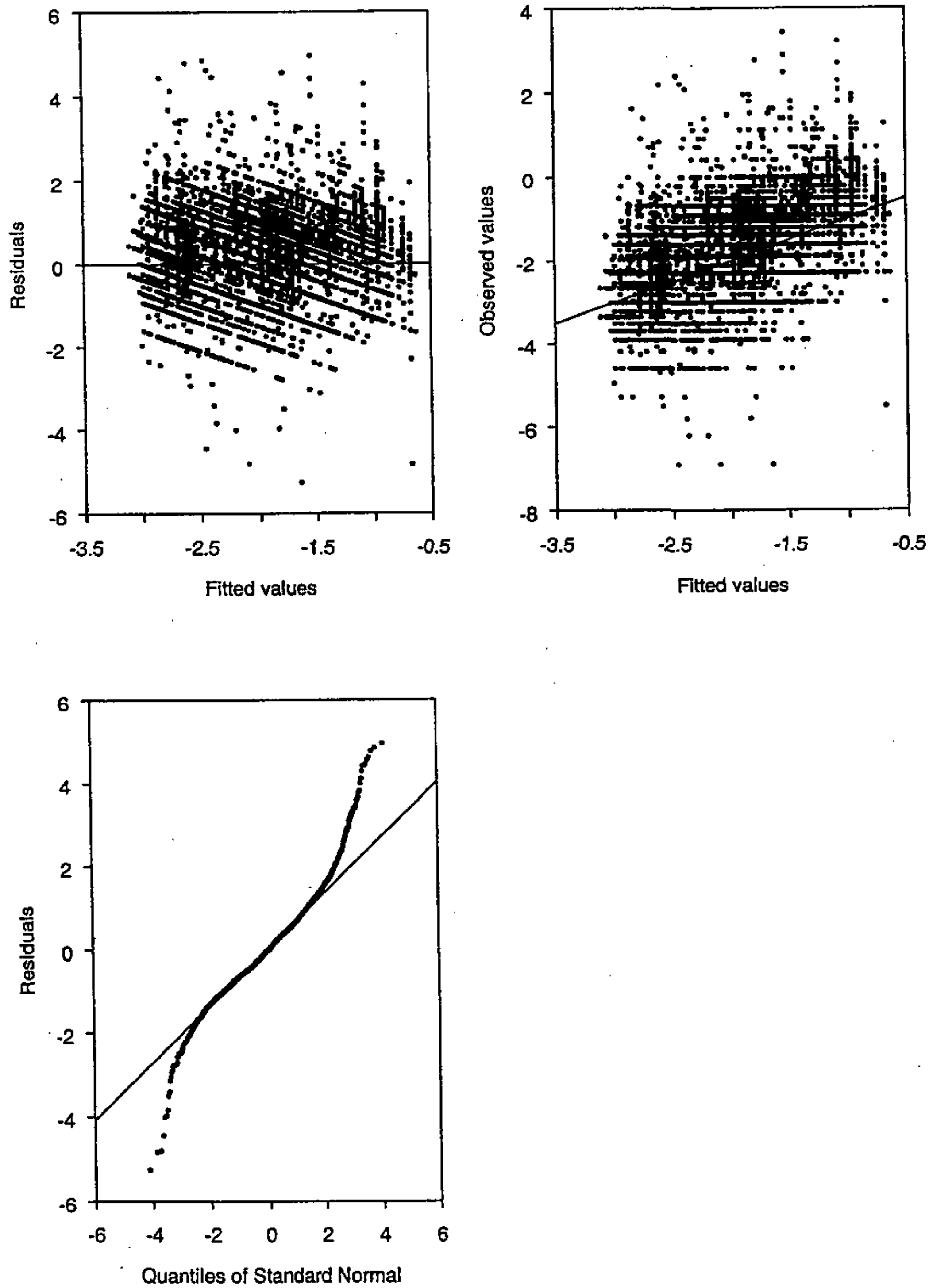

Figure D1: Diagnostic plots (log scale) of the lognormal model for the Chatham Rise fishery (excluding Statistical Area 404). Top left, fitted values versus residuals; top right, fitted values versus observed values; bottom left, normal quartile-quartile plot of the residuals.

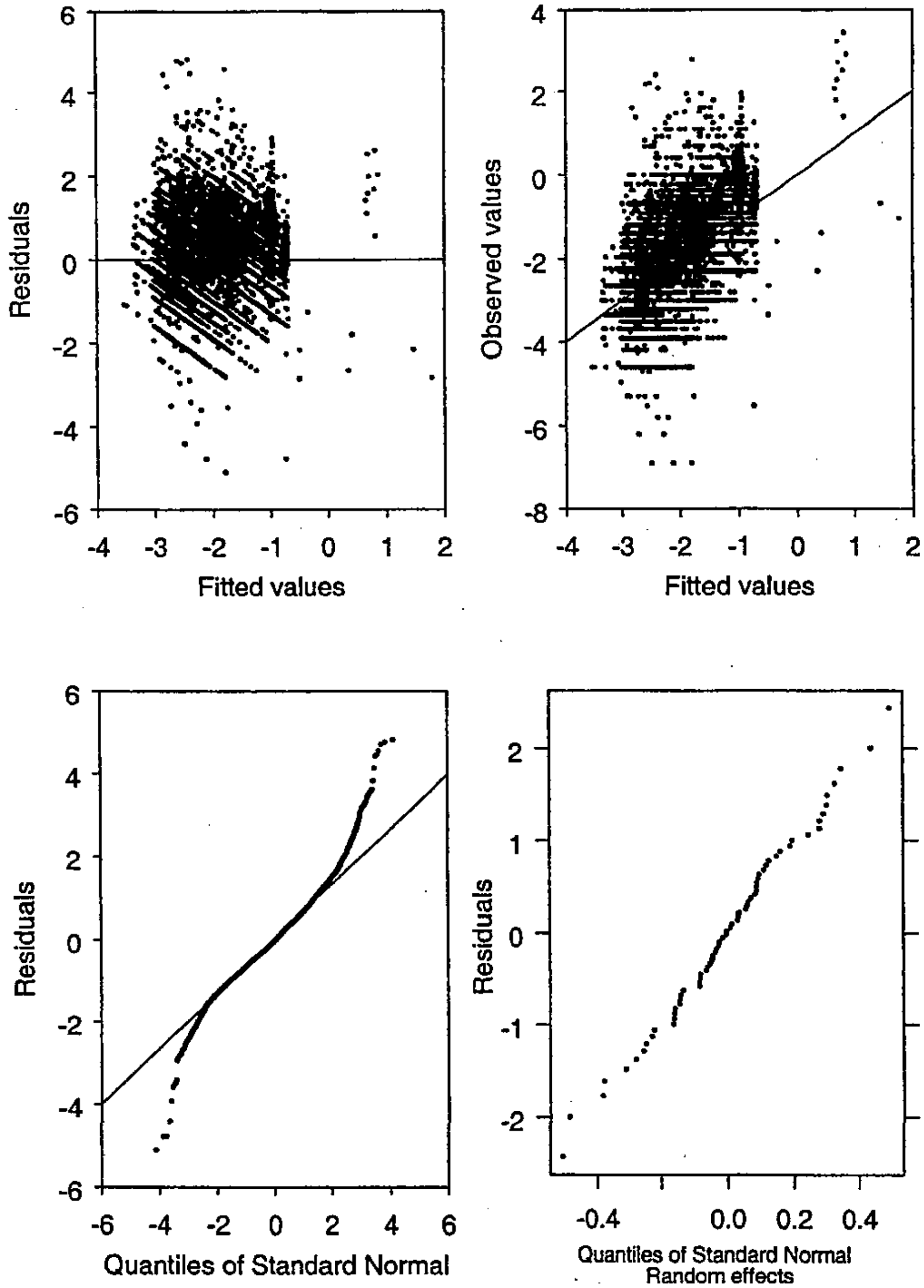

Figure D2: Diagnostic plots (log scale) of the mixed-effect model for the Chatham Rise fishery (excluding Statistical Area 404). Top left, fitted values versus residuals; top right, fitted values versus observed values; bottom left, normal quartile-quartile plot of the residuals; bottom right, a normal quantile-quantile plot of the residuals for method nested in vessel.

Appendix E: Diagnostic plots, Statistical Area 404 fishery

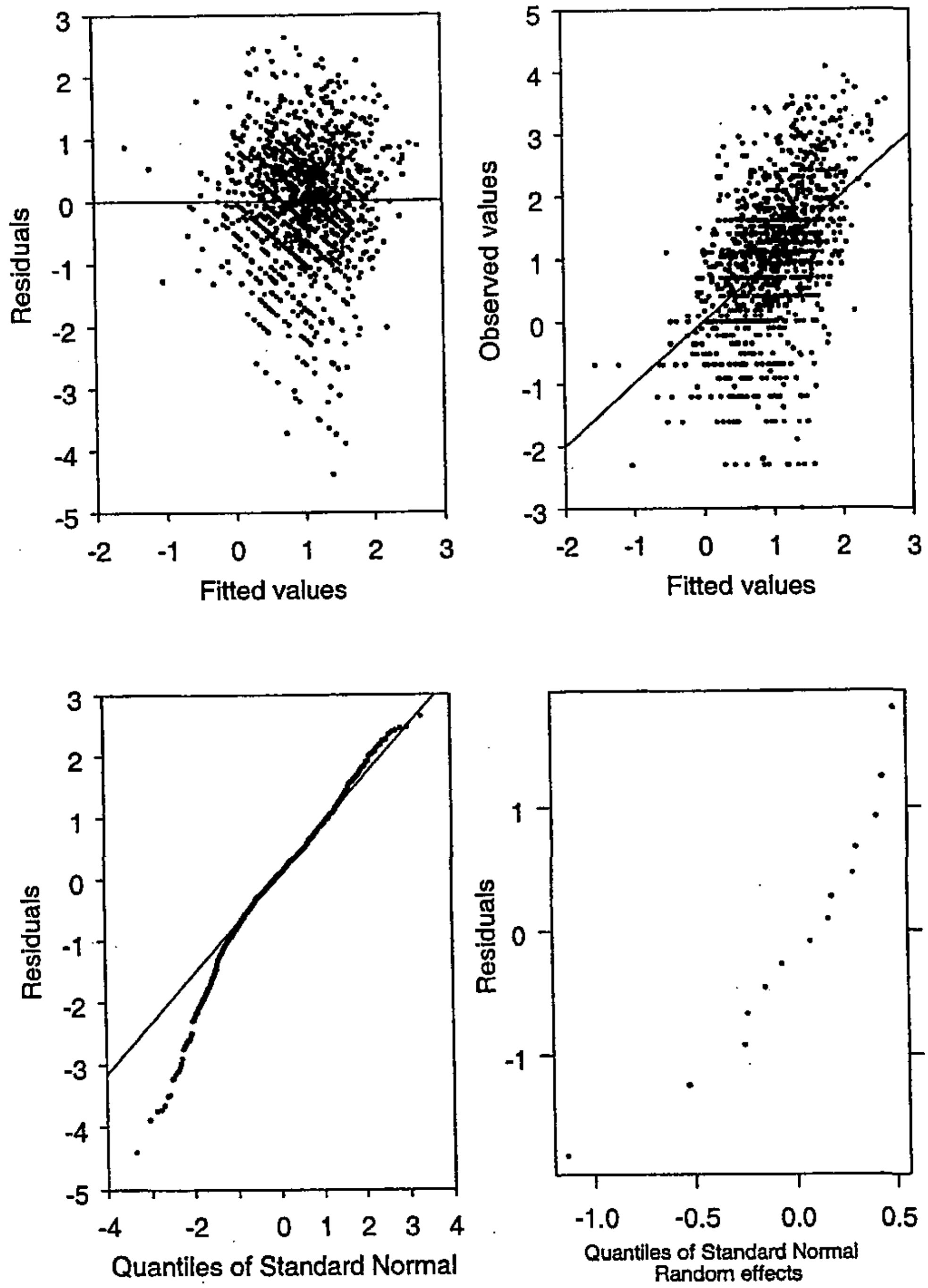

Figure E1: Diagnostic plots (log scale) of the fixed effects of mixed-effect model for the Statistical Area 404 fishery. Top left, fitted values versus residuals; top right, fitted values versus observed values; bottom left, normal quartile-quartile plot of the residuals; bottom right, a normal quartile-quartile plot of the residuals for method nested in vessel.

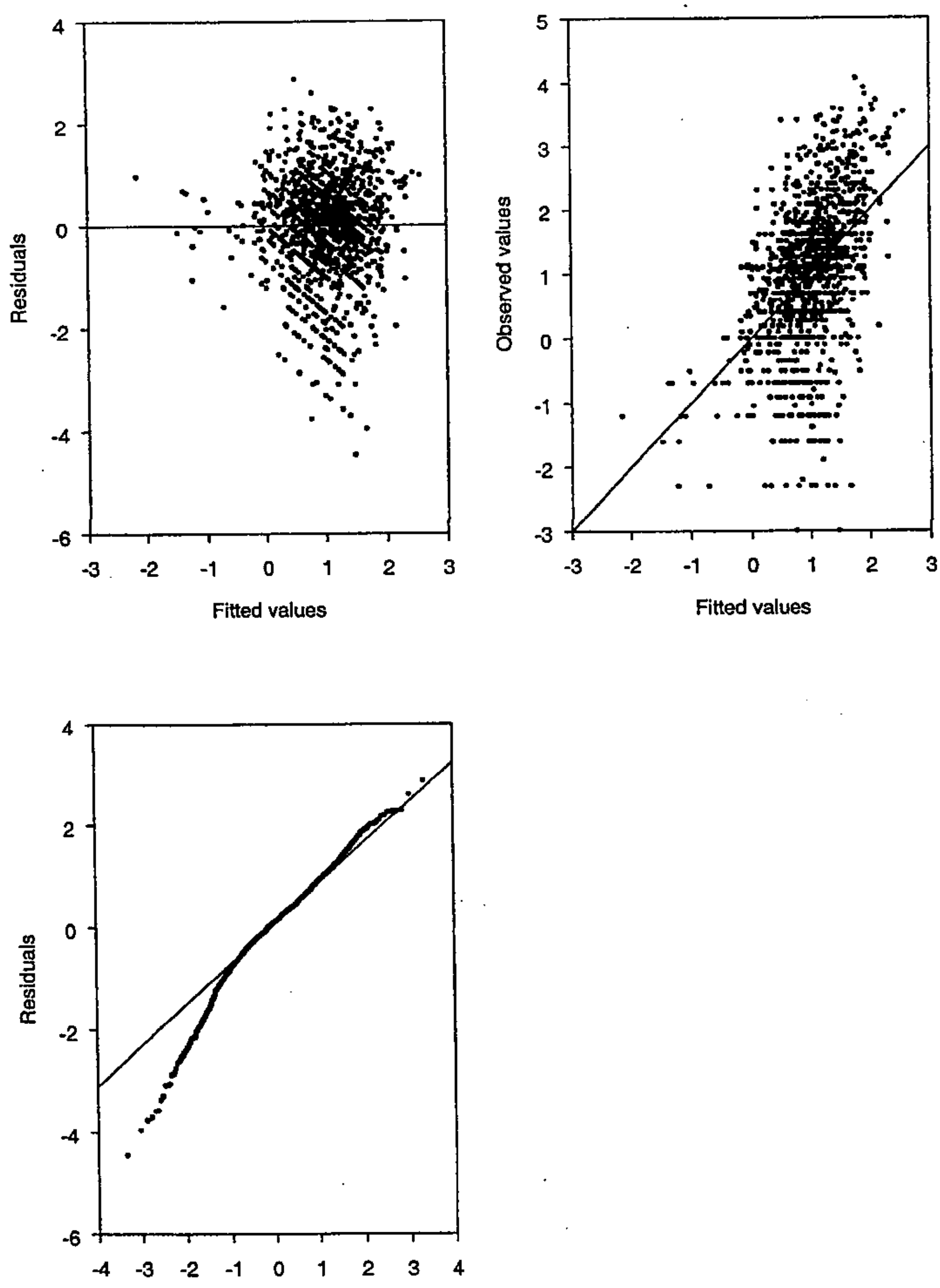

Figure E2: Diagnostic plots (log scale) of the lognormal model for the Statistical Area 404 fishery. Top left, fitted values versus residuals; top right, fitted values versus observed values; bottom left, normal quartile-quartile plot of the residuals.

Appendix F: Diagnostic Plots, Sub-Antarctic fishery

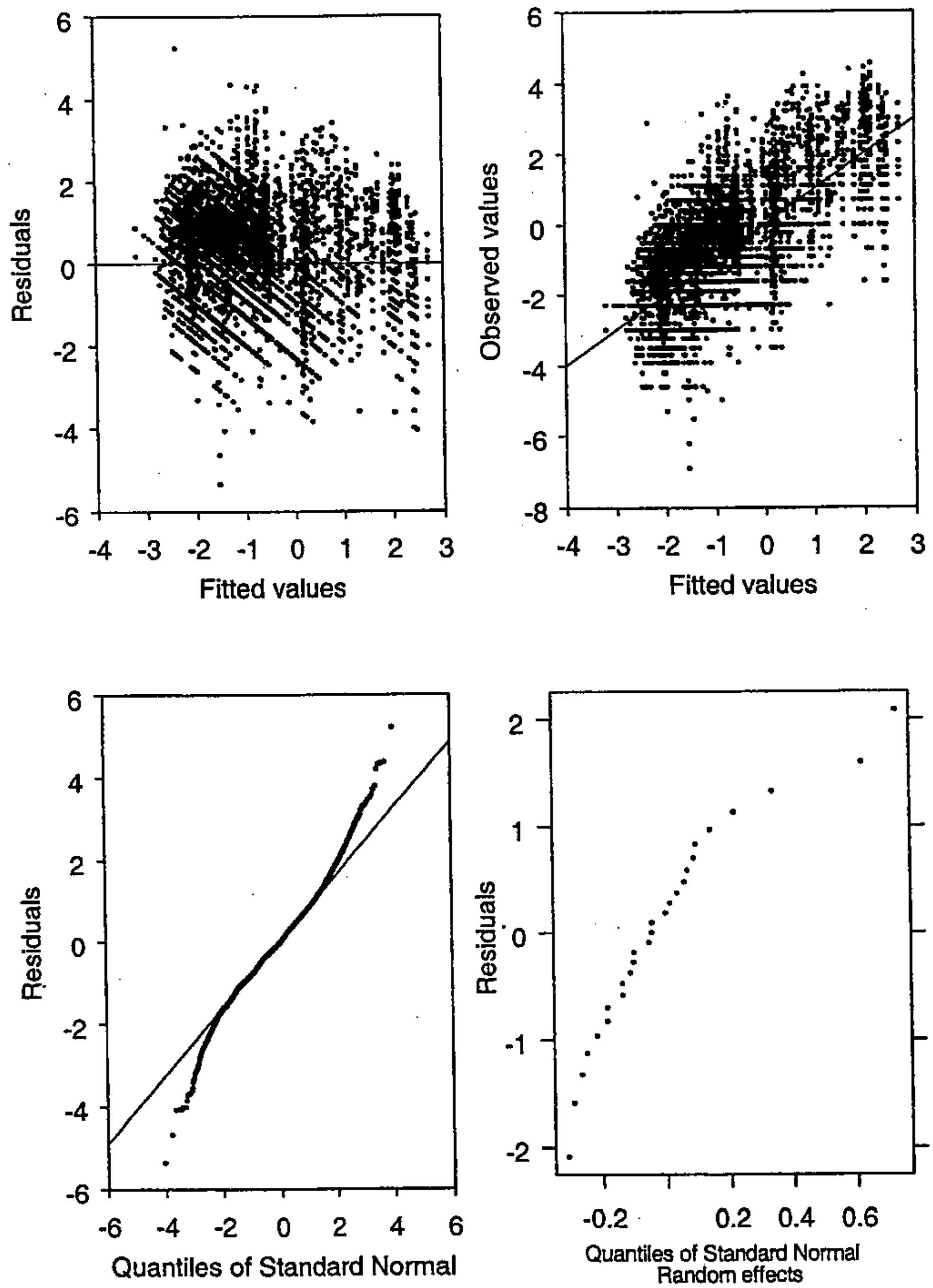

Figure F1: Diagnostic plots (log scale) of the mixed-effect model for the Sub-Antarctic fishery. Top left, fitted values versus residuals; top right, fitted values versus observed values; bottom left, normal quantile-quantile plot of the residuals; bottom right, a normal quartile-quartile plot of the residuals for the random effects (method nested in vessel).

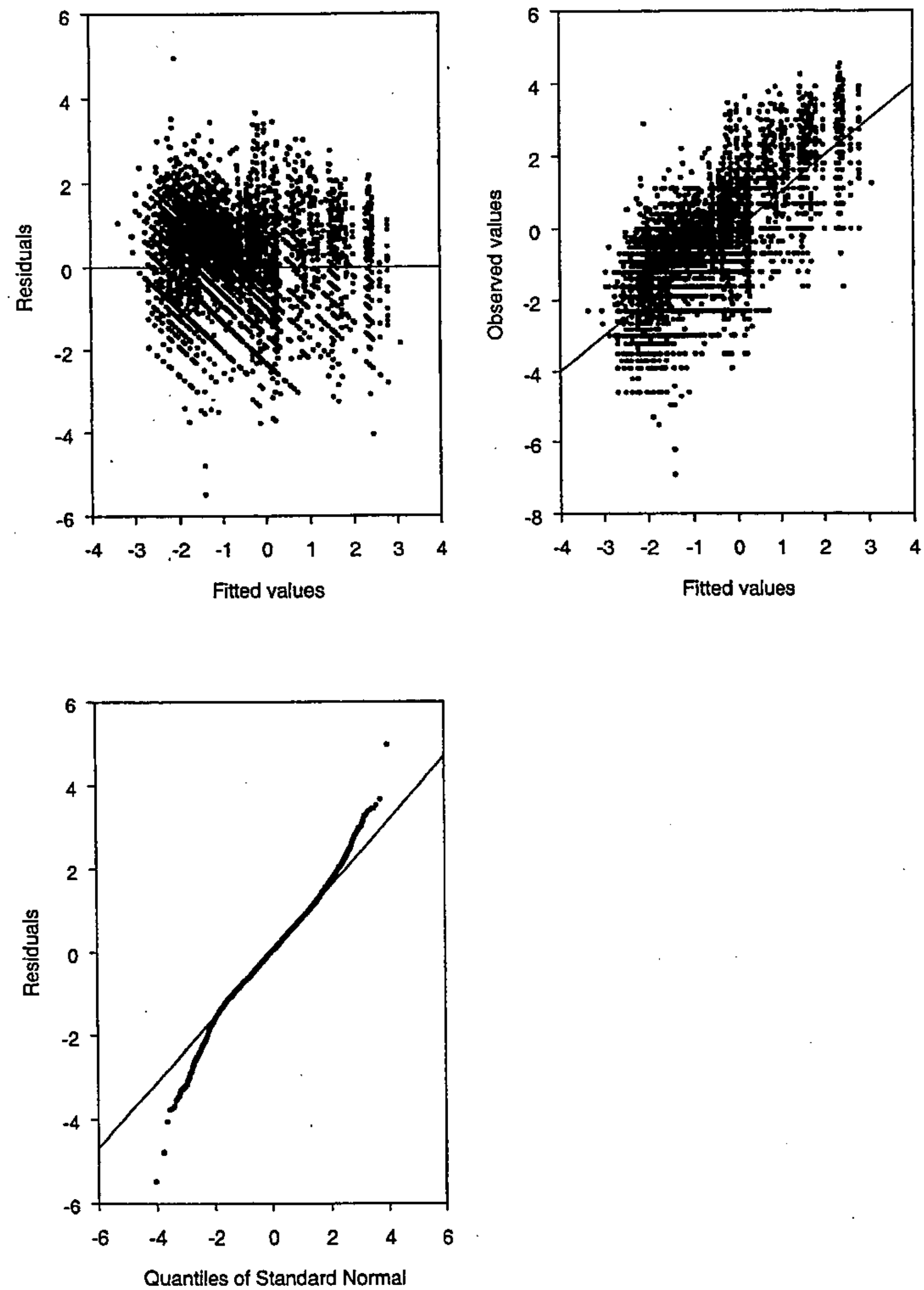

Figure F2: Diagnostic plots (log scale) of the lognormal model for the Sub-Antarctic fishery. Top left, fitted values versus residuals; top right, fitted values versus observed values; bottom left, normal quartile-quartile plot of the residuals.

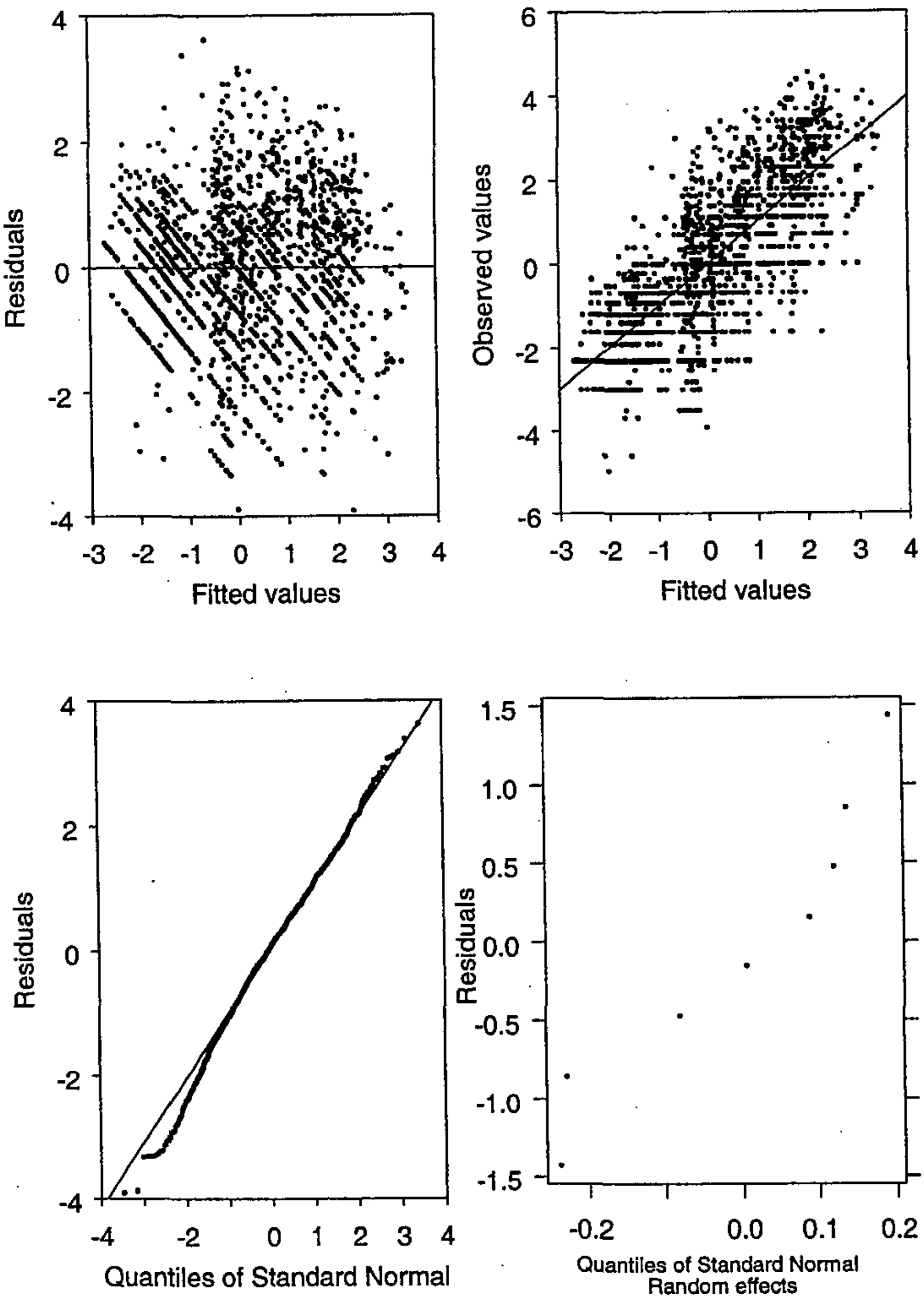

Figure F3: Diagnostic plots (log scale) of the fixed effects of mixed-effect model for the Stewart/Snares fishery. Top left, fitted values versus residuals; top right, fitted values versus observed values; bottom left, normal quantile-quantile plot of the residuals; bottom right, a normal quartile-quartile plot of the residuals for random effects (method nested in vessel).

Appendix G: Comparison of hake CPUE indices with trawl survey biomass indices

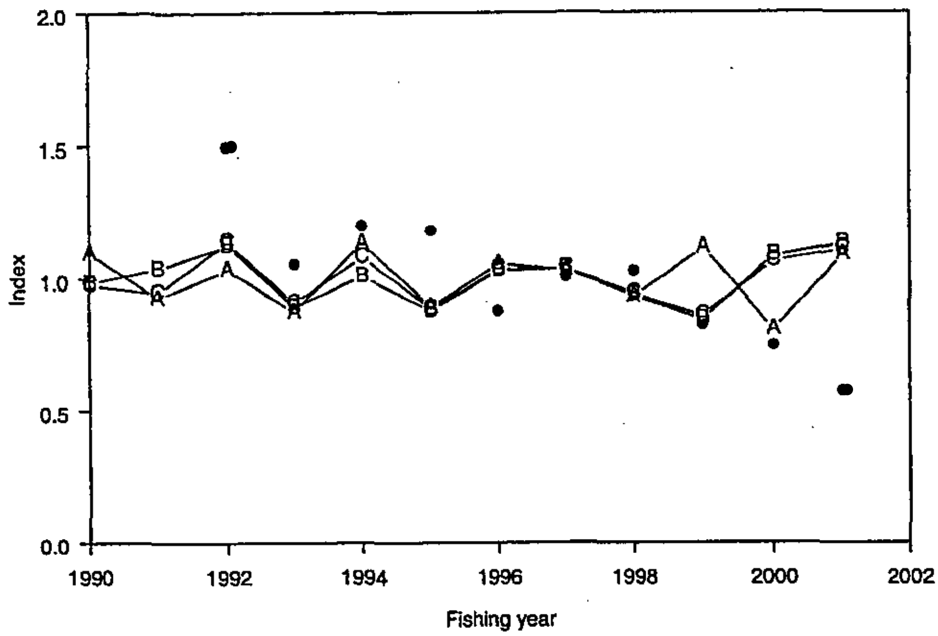

Figure G1: Comparison of trawl survey hake biomass indices (points) and CPUE indices from the Chatham Rise fishery (including Statistical Area 404) derived from the mixed-effect model ("A"), Chatham Rise fishery (excluding Statistical Area 404) derived from the mixed-effect model ("B") and the lognormal model ("C"). The indices have been scaled to a mean of 1.

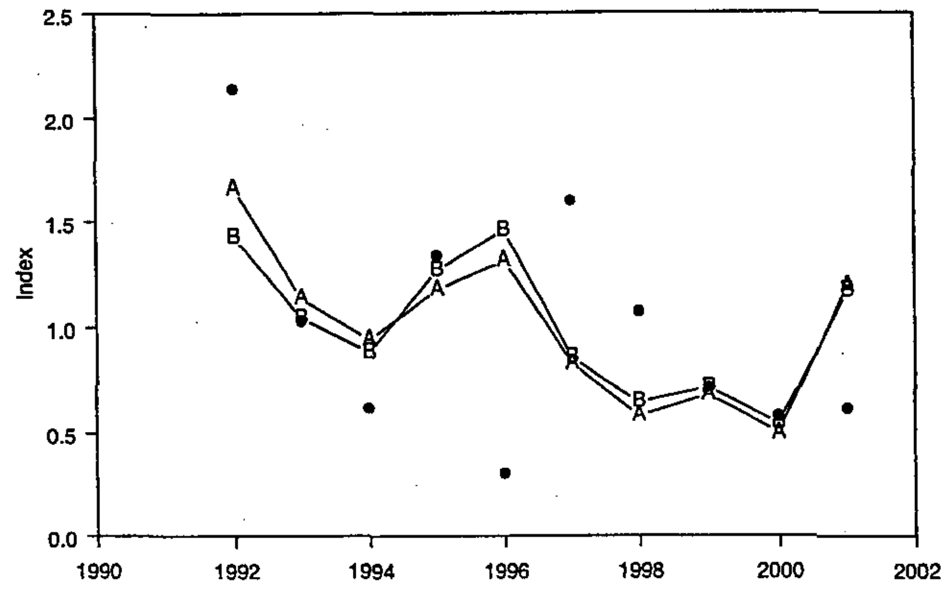

Figure G2: Comparison of trawl survey hake biomass indices (points) and CPUE indices from the Statistical Area 404 fishery derived from the lognormal model ("A"), and derived from the mixed-effect model ("B"). The indices have been scaled to a mean of 1.

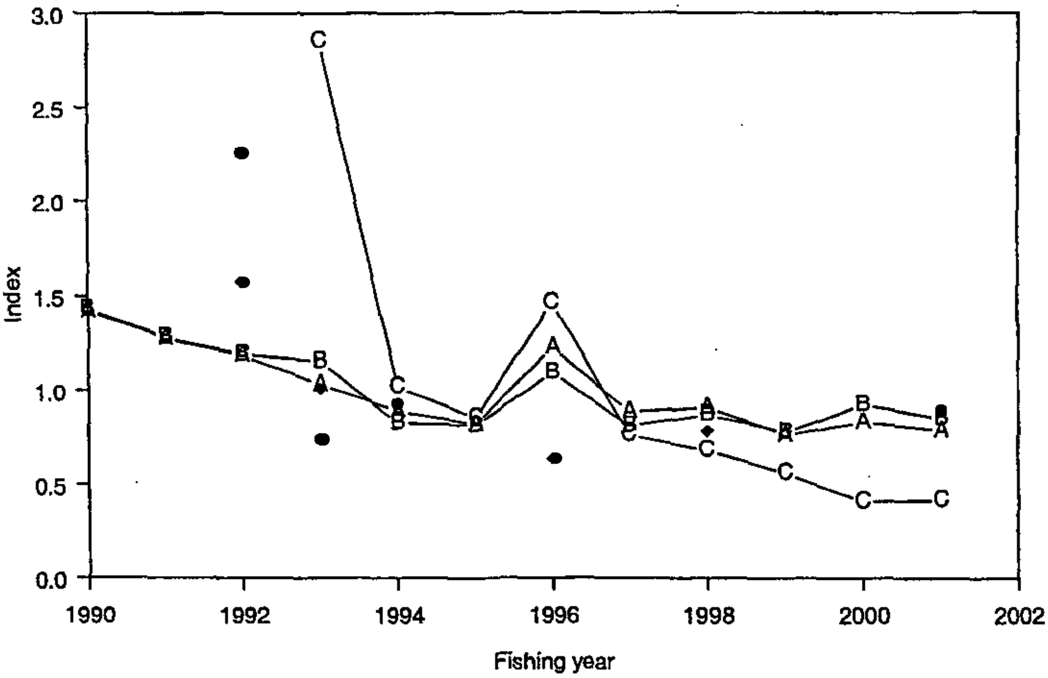

Figure G3: Comparison of trawl survey biomass indices (summer series, points; autumn series, diamonds) and CPUE indices from the *Sub-Antarctic fishery derived from* the lognormal model ("A"), from the mixed-effect model ("B") and the Snares fishery using the mixed-effect model ("C"). Note, the indices have been standardised with a mean of 1.